ISSN 1742-5867

Stammering Research

Editor Peter Howell PhD FIOA
Professor of Experimental Psychology
University College London

The British Stammering Association

The British Stammering Association
15 Old Ford Road
London E2 9PJ

Tel: 020 8983 1003
Fax: 020 8983 3591
Helpline: 0845 603 2001

www.stammering.org

Reg Charity No 1089967
Reg Company No 4297778

Stammering Research

A Journal Published by the British Stammering Association

Editor
Peter Howell
University College London
Gower St.,
London WC1E 6BT
England
Email p.howell@ucl.ac.uk

Associate Editors

Rosemarie Hayhow MSc. MRCSLT
Research Speech and Language Therapist
Speech & Language Therapy Research Unit
North Bristol NHS Trust
Frenchay Hospital
Bristol BS16 1LE
0117 918 6529
Email rosemarie@speech-therapy.org.uk

Dr Robin Lickley
Speech and Language Sciences
School of Social Sciences, Media and Communication
Faculty of Health and Social Sciences
Queen Margaret University College
Edinburgh EH12 8TS
0131 317 3686
Email
Rlickley@QMUC.ac.uk

Dr Trudy Stewart
Speech & Language Therapy Dept
St James Hospital
Beckett St
Leeds
LS9 7TF
Tel No: 0113 2064495
Email:
trudy.stewart@nhs.net

Notice

The British Stammering Association is a UK-based charity which seeks to promote understanding into the causes, treatment and understanding of stammering. Its activities include research into stammering which it supports through its vacation studentship scheme (http://www.stammering.org/research_schol.html) and the publication of Stammering Research (provided free of charge to all-comers).

Stammering Research is intended to promote public understanding of high quality scientific research into stammering and allied areas

If individuals wish to make a donation to support either of these initiatives, they should forward a cheque (payable to the British Stammering Association) to The British Stammering Association, 15 Old Ford Road, London E2 9PJ, or call the BSA on 020 8983 1003 (+44 20 8983 1003 from abroad) with their credit card details. If they wish this to be used specifically for either the vacation studentship scheme or Stammering Research, they should mark it accordingly on the back of the cheque. For information on tax-effective ways to support the charity's research activities, please go to http://www.stammering.org/donations.html.

Donors will be listed in the last issue of the appropriate volume of the journal unless they indicate otherwise. Companies wishing to make a donation or who wish to make enquiries about advertising in Stammering Research should address correspondence to Norbert Lieckfeldt at nl@stammering.org.

‘Stammering Research’.
An on-line journal published by the British Stammering Association
ISSN 1742-5867

Description

Stammering Research is an international journal published in electronic format. Currently it appears as four quarterly issues per volume (officially published March 31st, June 30th, September 30th and December 31st). The first issue of volume one will appear March 31st 2004. The journal is dedicated to the furtherance of research into stammering, and is published under the auspices of the British Stammering Association. It seeks reports of significant pieces of work on stammering and allied areas, such as other speech disorders and disfluency in the spontaneous speech of fluent speakers. The articles will include (though not be limited to) reviews in an area in which the author has produced eminent work and attempts to introduce new techniques into studies in the field. The journal will offer an opportunity to table topics where there are grounds for considering a major rethink is required, as well as detailing development and assessment of research-based techniques for diagnosis and treatment of the disorder. Submissions are encouraged that facilitate open access to scientific materials and tools. Articles are peer-reviewed, the role of reviewers being to ensure that accepted standards of scientific reporting are met, including correction of factual errors. Disagreements about interpretation of findings raised by reviewers will be passed on by the editorial board to the authors of accepted papers. These disagreements will not necessarily preclude publication of the article if they are judged to be topics that are suitable for open peer commentary. Once accepted, commentaries will be sought (actively and by self-nomination) from specialists within the field of communication disorder and its allied disciplines. These commentaries will be reviewed for style and content. The author’s responses will be reviewed in the same way. The article, open peer commentaries and author’s responses will be published simultaneously. Authors should contact the editor in the first instance with a short description of the topic area so that its general suitability can be assessed before full submission. Notification that a topic is suitable does not imply that the paper that is subsequently submitted will be accepted. Decisions about suitability will be made by the editorial board.

Editor
Peter Howell
Department of Psychology
University College London
Gower St.,
London WC1E 6BT
England

SUMMARY OF STEP BY STEP PROCEDURE FOR AUTHORING AN ARTICLE TO STAMMERING RESEARCH

1. Contact the editor with a brief outline of the proposed article. The editor and other board members make initial decisions only as to the suitability of the general area proposed. The primary function in this step is to ensure the topic is of sufficiently broad interest for, and within the remit of, the readership of Stammering Research. The intent behind this initial contact is to ensure authors do not spend time preparing articles on unsuitable topics. Review, empirical and theoretical work are all appropriate. Authors will be informed whether the judgement is that the proposed topic has a suitable, or too narrow, a focus. Indication that the scope is too narrow does not imply anything about the scientific standard of the proposed work. Neither does notification that a topic is suitable indicate that the submitted work will necessarily be accepted for publication (all submitted material has to go through the normal processes of peer review).
2. Submitted articles are peer reviewed in the normal way and an indication as to suitability of publication or not (possibly after revision) is notified to the author by the editor.
3. After an article has been accepted, the author cannot change the article. It is then made available for open peer commentary. Details how the accepted article can be accessed are posted on the British Stammering Association's website (www.stammering.org). Indications that the article is available for access are posted on http://www.mankato.msus.edu/dept/comdis/kuster/Internet/Listserv.html for ASHA members, the British Stammering Association's website (http://www.stammering.org, the stut-l list (stutt-l@listmail. temple.edu), the stutt-x mailing list (stutt-x@asu.edu), and on the stuttering home page (www.stutteringhomepage.com). The primary function in posting details about access available to an accepted article, is to alert potential commentators. A list of commentators is being drawn up and individuals are encouraged to submit their nominations (for themselves or others).
4. See the next page for precise details how to prepare a commentary and the timetable allowed for this. When preparing a commentary, authors might find it helpful to consult a recent issue of Stammering Research to see the range of comments that are appropriate, the style and format of commentary submissions.
5. All accepted commentaries are available to the author of a target article from receipt until two weeks invitations for commentaries has closed. In this time, the author can prepare a response to commentaries. The response will be peer-reviewed by the editorial board. Further details are given on the next page and authors should again consult a recent issue of Stammering Research to see the sorts of comments that are appropriate, style and formatting of a submission.
6. On completion of this process, the target article, commentaries and response to commentaries will be published together in the next issue of Stammering Research. Authors are responsible for preparing their articles according to the stipulated format. The current and previous issues of the journal are available as PDF files at http://www.speech.psychol.ucl.ac.uk/.

Notes about commentaries for Stammering Research
ISSN 1742-5867

Once a manuscript has been accepted as a target article, the authors cannot change it. The manuscript needs to be available for commentary before it is officially published so that commentaries and the author's responses can appear simultaneously.

Manuscripts are posted for commentary on http://www.psychol.ucl.ac.uk/ under *Stammering Research*. Commentators are alerted as indicated on the previous page.

Manuscripts will be available for peer commentary for six weeks. Commentaries have to reach the editor, or associate editor, responsible for the article within that time (late submissions will not be accepted). Commentaries should ordinarily not exceed a total (including references and other material) of 1,000 words. The commentaries have to conform to APA style conventions.

Commentaries should be sent by email as soon as possible within the six-week period the article is open for peer commentary. The commentary should appear within the body of the email text (not as an attachment) and be sent to psychol-stammer@ucl.ac.uk. Authors of target articles will receive commentaries as they are accepted and have two weeks from close of submission of commentaries to complete their responses.

Commentaries will be peer-reviewed and edited for style as well as content. Authors of commentaries need to establish the relevance of their submission to the target article at the outset, and preferably also show an awareness of the wider work of the target article's author.

If there are several commentaries which raise the same point, the editorial board reserves the right to group them together and prepare them as a single coauthored commentary. In this (probably rare) eventuality, the authors will have the opportunity to see the manuscript and decide whether they wish to be included on the list of authors.

Editing and revision of commentaries will be completed within two weeks of close of submission. Revisions that are not satisfactorily completed in this period, or that are received late, will not be published.

In exceptional circumstances, new commentaries may be considered as submissions for on-going commentaries that will appear in later issues of Stammering Research. These will be treated in the same way as initial commentaries (e.g. in terms of target authors responses).

Formatting Accepted Publications in Stammering Research

Peter Howell[1], John Smith[2], and John Doe[3]
[1]*Department of Psychology, University College London, Gower St., London WC1E 6BT England*
P. Howell@ucl.ac.uk

[2]*Stuttering Treatment Clinic, Somewhere, Some Country*
Smith@some.email.address

[3]*For private citizens, house number and street, city and postcode/ZIPcode, Country*
Doe@email.address.if.you.have.one

Abstract. A short abstract summarizing the significant content and contribution of the paper should be included here. This page illustrates and describes the format for paper submissions. Authors are requested to adhere as closely as possible to this format once an article is accepted. The abstract should be in Times New Roman 9-point font, justified with left and right margins indented 1 cm in from the margins of the main text.

1. Introduction

Articles and commentaries should initially be submitted in APA format. After an article or commentary is accepted, it needs to be prepared according to the journal format as indicated next. Articles and commentaries must be in Word format. An article will typically be up to **15,000 words**. A commentary should preferably be up to **1,000 words**. Authors may submit longer articles or commentaries for consideration but these may be reduced in length by the editor. Articles with fewer than 15,000 words and commentaries with fewer than 1,000 words are acceptable if the author can demonstrate sufficient content and contribution. Typically commentaries will have an abstract, usually only a single section in the text headed so as to identify the target article, and will not use diagrams or photographs. However, if an author needs to use more than one section heading and diagrams or figures, then they should follow the same instructions as for preparation of a target article. Each page of an article should consist of single column, of single-spaced text in a 16cm x 24cm column using **A4** or **US Letter** settings on your word processor as illustrated in Figures 1 and 2. Figures should be numbered consecutively and appear close to the text where they are mentioned.

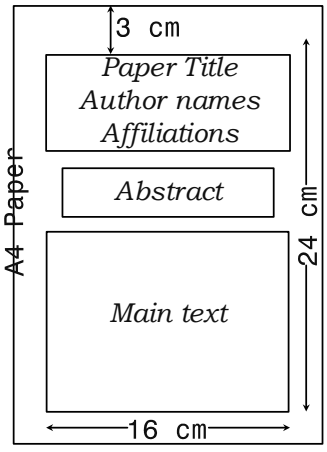


Figure 1: First page format

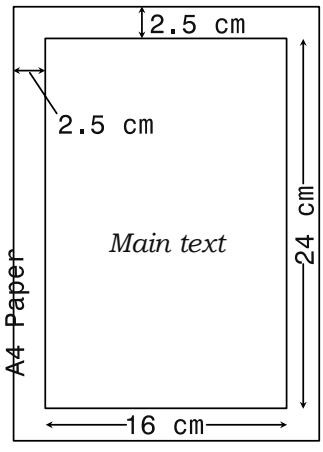


Figure 2: Subsequent page format

2. Detail of styles

The article or commentary title should be bold and centred using 14 point Times New Roman font. Authors' names, affiliations and email details should be centred using 10 point Times New Roman font. The author's affiliation should be italicized. The main text and the bibliographical references must be justified and single line spaced. The main text should be in 10 point Times New Roman font with numbered section headings in 11 point bold font.

All references should be cited using APA referencing styles. For example a publication which is referred to as support for a statement would be cited in the text this way (Howell & Sackin, 2002) whatever the number of authors. When an article is referred to directly in the text as in "… in the work of Howell and Sackin (2002) the …" only the year is placed in brackets. If there is more than one reference from the same authors in the same year then they are distinguished by using different letter designations after the year as in 1996a, 1996b etc. In the references below, examples are given of how a conference paper, a journal paper and a book would be listed. All references should be listed at the end of the paper using 9 point Times New Roman font.

All figures, and diagrams must be good quality black and white images suitable for readers to display and print. Colour illustrations or text can be used, but bear in mind readers who want to print articles may not have access to a colour printer. When an article is accepted, figures and pictures must be inserted in the word file in the exact position they will appear in the publication. Any format for figures, pictures and diagrams may be used provided they allow good quality reproduction for readers who wish to print off a copy.

References

Howell, P. (2002). The EXPLAN theory of fluency control applied to the treatment of stuttering by altered feedback and operant procedures. In E. Fava (Ed.), *Pathology and therapy of speech disorders (pp. 95 -118). Amsterdam: John Benjamins.*

Howell, P., & Sackin, S. (2002). Timing interference to speech in altered listening conditions. *Journal of the Acoustical Society of America*, *111*, 2842-2852.

Rosen, S., & Howell, P. (1991). *Signals and Systems for Speech and Hearing*. London and San Diego: Academic Press.

Stammering Research

A Journal Published by the British Stammering Association

Volume 1, Issue 2, July 2004

Contents

RESPONSE TO COMMENTARIES

RESEARCH DATA, SOFTWARE AND ANALYSIS SECTION

Editorial for Stammering Research

As stated in the editorial to the first issue of the journal Stammering Research, the journal is dedicated to dissemination of a wide spectrum of opinion on topics in this field of research. In the first issue, target articles on specific topics were published along with open peer commentaries and responses by the original authors. In this issue there are two further target articles, commentaries and responses to commentaries.

Also in this issue, Stammering Research starts a new initiative that attempts to extend involvement in research into stammering (the 'Research data, software and analysis' section). This section is intended to provide; a) access to facilities in order to make it easier for individuals already doing research in the area, b) resources that will allow people not previously involved to examine stammering scientifically, c) open access to teaching materials (such as samples of speech data from speakers who stammer), d) data that can be reanalyzed by different techniques, e) rare data so that, when other instances are pooled, a big enough cohort can be built up for their analysis, f) open access outcome data on treatments so that they can be used to support evidence-based practice.

It is anticipated that diverse forms of data will be made available through this section of the journal. Audio data are the first that are released. Details about how to access speech samples obtained from speakers who stammer in spontaneous monologue are provided, pointers to a wide range of sources for the analysis of these data and some illustrative analyses with one free software package are described. These data can be used for teaching how stammering is assessed. Reports of original and comparative analyses of these data have been invited as submissions for Stammering Research. Later issues will provide samples of data under frequency-shifted and delayed-auditory feedback manipulations (Howell, 2004). These will allow readers to assess the influences of these manipulations (based on the numerous inquiries I receive, this should be of particular interest to people who stammer) and also, researchers can use these data to establish what changes in vocal control occur when speakers become fluent under these manipulations. Similar data using other techniques are welcomed as submissions (these need to be 1) of general interest, 2) collected under ethically appropriate conditions with permission, 3) consistent with data protection legislation and 4) of a sufficient quality for analysis). It is also planned to provide imaging and articulatory data for reanalysis. This should sidestep the problem that people cannot get involved in research with these and other techniques because the data are costly or impossible for individual researchers to obtain. Examples of rare data that it is intended to post are cases of bilingual speakers who stammer, data where pooling is necessary because individual researchers have too little data to meet statistical power considerations (e.g. family history data, Felsenfield, 1998) and so on.

As is demonstrated by the Howell and Huckvale (2004) publication, the ethical-legal considerations about the audio data that are supplied are surmountable, as they are with genetic, articulatory and imaging data that have been obtained from a subset of these speakers. Technical issues do strain resources (e.g. raw imaging data require extensive storage space) so, in the short term, it might only be possible to provide processed data (e.g. brain activation regions). Technical advances should change this situation soon.

I believe that Stammering Research makes goals a)-f) easier as i) it is a not-for-profit journal and ii) it is web-based. i) means that the cost of providing data does not have to be judged against the earning potential this resource offers. It does rely on support from technical staff, the speakers and research personnel who gave generously of their time. Stevie Sackin and latterly Steve Davis have played vital roles in recording and archiving the audio tapes. Jon Bartrip also deserves particular mention as he has been instrumental in establishing the data handling facilities and will continue this role to allow the many, and extensive, forms of data we anticipate Stammering Research to supply in the future.

References

Felsenfield, S. (1998). What can genetics research tell us about stuttering treatment issues? In A. K. Cordes and R. J. Ingham (Eds.), *Treatment efficacy for stuttering: A search for empirical bases. San Diego: Singular Publishing Group Inc*.

Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Research*, *1*, 31-46.

Howell, P. & Huckvale, M. (2004). Facilities to assist people to research into stammered speech. *Stammering Research*, *1*, 130-242.

Peter Howell
June 2004

TARGET ARTICLE

Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?

C. Savage[1] and E. Lieven[2]
[1] *Department of Psychology, University College London, Gower St., London WC1E 6BT, England*
c.savage@ucl.ac.uk
[2] *Max Planck Institute for Evolutionary Anthropology, Leipzig*
lieven@eva.mpg.de

Abstract. The usage-based approach to language development suggests that children initially build up their language through very concrete constructions based around individual words or frames on the basis of the speech they hear and use. These constructions gradually become more general and more abstract during the third and fourth year of life. We outline this approach and suggest that it may be applied to problems of fluency control in early child language development.
Keywords: Usage based approach, generativist approach, EXPLAN.

1. Setting the scene

A child begins to control the articulators in an effort to make vocal sounds shortly after birth. These early efforts provide the basis for most language development which develops over the course of three years from single words to simple, and subsequently more complex, grammatical utterances. Any difficulties that a child experiences as she tackles the acquisition of each component skill during language development may result in different types of fluency problem. During the vocabulary spurt at around 18 months, for instance, a coincident increase in naming errors also occurs. This is characterized by perseveration of recently produced words. As well as reflecting the vulnerability of newly acquired items to retrieval error (Gershkoff-Stowe, 2002), it has also been suggested that the errors could reflect the early system-wide fragility of the word retrieval process (e.g. Marchman & Bates, 1994; Plunkett & Marchman, 1993). Over time, with increased practice, these naming errors disappear, suggesting that the demands on the language production system become better aligned with its capacity.

Other problems that occur during the development of fluency control may be more persistent and it is essential to know why. Here we will be examining, from a usage-based approach, what problems in early language development could lie behind stuttering. Stuttering is a disorder that is particularly prevalent in childhood; its modal onset age is 3 years (past the one-word stage of language development studied by Gershkoff-Stowe, 2002). It is possible that the fluency problems experienced at 3 years are a 'side effect' of the child acquiring another language skill around this age, similar to the disfluency that results during rapid lexical development. One possible explanation is that grammar is starting to become more general and abstract. Increased disfluency could result from the child's more sophisticated attempts to combine words into grammatical forms (Bernstein-Ratner, 1997). From this perspective, the high rate of spontaneous recovery during childhood (Andrews, Craig, Feyer, Hoddinott, Howie and Neilson, 1983) could be due to 'realignment' of language demands with the child's capacity to meet them during later language development. The hypothesis is, then, that stuttering prevails at points of vulnerability in language development when the system is under strain through an advance in acquisition of a linguistic skill. However, care must be taken to specify what is meant by 'grammatical development' as the particular definition will have ramifications for when this skill is thought to onset. Children combine words from as early as 2 years. They are able to produce novel utterances on a principled but limited basis by, for example, slotting certain nouns or verbs into low-scope frames. This almost always results in utterances in canonical word order, albeit often with material missing (e.g. Lieven, Behrens, Speares and Tomasello, 2003). The nature of children's early grammar will be discussed in more detail later in this article. The important issue is that, given that children gradually develop their ability to combine words grammatically from before 3 years, why do the problems in speech control start relatively late (i.e. at 3, rather than 2, years)? The current article will provide a developmental perspective on stuttering and speech production in general. The nature of children's early linguistic knowledge will be discussed (section 2), and how this knowledge relates to online speech production (section 3), the mechanisms of speech production and how disfluencies could arise are

discussed in section 4. A specific theory about the development of stuttering (EXPLAN) is reviewed in section 5, and how this theory might relate to the language development models reviewed in section 2 (section 6). Finally future directions of research are proposed that would take into account the findings from all these fields in a unified way (section 7).

2. The Nature of Early Linguistic Representations and its Relationship with Language Production

The modal onset age of stuttering is 3 years yet syntactic development precedes this and has progressed beyond the one word stage. The reason that age of stuttering onset does not coincide with age of language onset may have important ramifications for the language acquisition debate. Stuttering begins during early language development, suggesting that an adequate explanation of stuttering must include a developmental perspective. The critical question here is how to characterise children's underlying linguistic competence across language development and how to measure it. Studying the abilities of young children places restrictions on the techniques we can use. As always in developmental research, much of the challenge for researchers lies in finding appropriate methodologies. Standard tests of language development are not widely available for children of this age and the usefulness of a gross measure such as mean length of utterance (MLU) is limited. MLU is often measured on naturalistic data but a single MLU measure on a small sample of speech is not sufficient to characterize a child's state of language development. MLU does not differentiate sufficiently between fine-grained distinctions of language sophistication. So, for example a high MLU score might be obtained for a child who strings together formulaic and phonologically simple units using "*and*" (example 1 below) compared to a child who produces shorter phrases but does so by using more abstract and creative production schemas (example 2). Thus the creativity and the length of an utterance do not necessarily concur.

1. *Mum laughed and Dad laughed and Bob laughed* (11 morphemes).
2. *I bloomed these flowers (= made these flowers bloom)* (6 morphemes).

These examples indicate that MLU is not ideal for examining factors that may affect fluency, such as syntax and phonological complexity. In this paper some of the methods used to study children's naturalistic language development will be explored in depth, with particular reference to the 'Usage-based Approach' (Langacker, 1987) (UB approach). We will examine the relationship between the existing work on fluent language development with that on developmental stuttering and we will suggest how the UB approach might provide new methods for studying stuttering near onset. First, however, we outline the potentially very different predictions about disfluency that stem from the UB approach and the traditional formal, approach to language development.

The generativist approach (starting with Chomsky's work in the 1960s) conceptualises linguistic competence as based on Universal Grammar (UG), and maintains that children have the abstract categories of UG from the start. In contrast, the UB approach suggests that children initially lack abstract knowledge and instead start their grammatical development by building up lexically-specific patterns. The basic problem with the generativist theories is that they analyse the child's language in terms of the categories of adult grammar, thus 'defining away the problem'. There are, however, a large number of studies that suggest that children's early language is precisely *not* that of the adult grammar (see sections 2.1. and 2.2. below). Interestingly, the two assumptions make different predictions for both the age of stuttering onset and the location of disfluencies in speech. The generativist approach predicts: (a) Children would stutter as soon as multiword utterances occur if the problem stems from syntax; (b) Children would stutter around phrase structure boundaries (e.g. before noun phrase, NPs, or verb phrases, VPs or prepositional phrases, PPs) because these are the points in utterances where speech is planned. The predictions of the UB approach are different: (a) Stuttering may start at a later age (around 3 years) because experimental evidence (see section 2.2.) suggests that children only start to be productive with the more abstract aspects of grammar e.g. the argument structure of verbs, towards the end of their third year; (b) If children stutter they will do so at the boundaries of psychologically real planning units that may not always coincide with those of traditional syntactic theory. Thus UB theory maintains that frequently used patterns will continue to be represented lexically into adulthood as well as being analysable in terms of abstract syntax (e.g. "*I dunno*" and "*I don't know*", Bybee and Scheibman, 1999). The implication of this is that stuttering could occur within a constituent if it exists as a representational unit (e.g. *I wanna*) as the child attempts to combine previously learned frames (*I wanna* with *biscuit* leading to "*I wanna bbbbbbiscuit*" instead of "*I want aaaaaa biscuit*", where the stuttering occurs at the NP phrase boundary).

The generativist approach to syntactic development does not predict the modal age of stuttering onset (3 years), whereas the UB approach may provide an explanation for this age of onset. Research within the UB approach suggests that, for normal development, 3 years is the stage in development

when grammar is being reorganised from lexically-specific schemas into more abstract syntax. Stuttering would then result as a 'side effect' of the increased demands that this reorganization places on the child's processing system. Later, the child's speech production system has adapted to perform these operations more efficiently, and the older child will achieve greater fluency. The proposal is a grammatical parallel of Gershkoff-Stowe's theory of how fluency is affected by the development of lexical access.

The question remains open as to why some children continue to stutter for longer than others and why a minority continue to do so into adulthood. There are processing theories that have begun to answer this question, which will be explored later. The UB approach may provide an original avenue for exploring differences between children who stutter and fluent speakers across development. For instance, are there differences in the size of the units that children who stutter store and employ relative to fluent children? Are children more likely to speak fluently if they use more concrete and directly accessed language units? How does the development over time of more abstract units relate to creativity and efficiency in language production? The same methods that have been applied to ascertain the nature of early syntactic knowledge in normal language development could be used to help answer similar questions for children who stutter. A discussion of the existing research for normally-developing children will make more concrete the second prediction made above about the location of disfluencies by revealing the planning units that young children use.

Observational studies of naturalistic child language data in the 1970s indicated that young children use at least some of their language in item-specific ways (e.g. Braine, 1976). Early attempts to explain the data proposed that young children arranged their linguistic knowledge around limited scope cognitive-semantic categories that they learned from their language input. Adult-like verb-object pairs appeared in the data but seemed to belong to a variety of independent positional patterns, such as "see + X", "want + X", and "have + X". Braine suggested that each pattern was associated with a specific semantic content, for example a pattern related to oral consumption, associated with *eat*, *bite*, and *drink*. The early cognitive-semantic theories were the first attempts to radically rethink the fundamentals of the language acquisition problem after the generativist tradition emerged. However, subsequent research revealed that children's use of grammatical structures crosses the boundaries suggested by semantic groupings (e.g. Levy, 1983; Valian, 1986; Pine, Lieven and Rowland, 1998). The UB approach has subsequently pioneered a more thorough approach to data that allows more accurate identification of the ways in which children's early grammatical knowledge is restricted. The approach adopts rigorous analytic methodologies that use consistent statistical criteria to classify structures within naturalistic data and backs these up with experimental data. Without this systematic treatment of the data there is a risk of either under-, or over-, estimating what a child can do (e.g. experiments that are too hard for the child to perform or naturalistic analyses that accept a structure as acquired at an abstract level on the basis of one exemplar).

2.1. Naturalistic Data

In working with naturalistic data there is always the problem of how to determine whether a structure is productive or rote-learned. Of course, without a 100% sample of what the child says and hears, this can never be done with full certainty for any utterance. But this is all the more reason to develop consistent definitions and to apply them uniformly. However the definition of how to classify a structure as having been acquired has not been consistent either within or between analyses. The number and type of instances that are observed in a corpus are interpreted differently depending on the theoretical framework within which an author is working. Radford (1990, 1995, 1996), for example, draws a distinction between 'acquisition' and 'mastery'. After acquisition but before mastery, children alternate between correct and incorrect use, and sometimes omit functors whilst learning exactly how to manipulate their newly acquired grammatical knowledge (Radford, 1990, 1995, 1996). Gathercole and Williams (1994) point out, though, that Radford applies the criteria non-uniformly. The same type of utterance is classified differently depending on which stage it appears in. For example he describes a 'How are you?' that occurs during his 'lexical stage' as rote learned because it is repeated monotonously in the same transcript and is the only example of a correct wh-question. In the same 'lexical stage', conversely, he classifies as productive sporadically produced questions such as 'Doing what?', which could just as well be rote learned from adult echo questions like 'You're doing what?'. Furthermore, his functional stage evidence for adjectival complements rests on only one example. A different example comes from Valian (1991) who based category assignment on the surrounding linguistic and social context of an utterance, and assigned a word in a child's utterance to the same category that it would be assigned to in adult speech. The basis for this judgment was adult grammar, which effectively rules out other possible interpretations. Valian's work assumed *a priori* the existence of the categories for which evidence is being sought, making the argument circular.

The UB approach tackles the problem of non-uniformity in data analysis by abandoning the traditional distinction between underlying competence and surface performance. The utterances of the child are instead treated as a direct indication of the child's underlying representations. The justification is that a canonical utterance produced by a child cannot indicate how abstract her underlying knowledge is. She may have produced it on the basis of an abstract generative grammar or she could just as easily have produced it directly from a string of words and phrases previously heard and learned. There is no absolute method of distinguishing between the two accounts in the case of naturalistic data because both routes would produce an adult-like utterance (either by knowing adult-like grammar or by hearing adult-like speech in the input).

The only evidence in naturalistic data from which we can conclusively infer abstract knowledge is the creative language use seen in examples of overgeneralization errors like the one given in (2) above and '*She giggled me*' (Bowerman, 1982). These have been systematically studied and reported for only a few children and they only occur with any frequency after 3 years of age with almost none before 2.5 years (Pinker, 1989). So what is the nature of children's representations before 3 years? After the early work on cognitive-semantic, item-based patterns in early language, more recent models have suggested that early utterances follow input-based lexical patterns (e.g. Braine, 1988; Pine and Lieven, 1993; Lieven, Pine and Baldwin, 1997; Pine, Lieven and Rowland, 1998; Tomasello, 1992; Tomasello, 2000). One of the first influential accounts in this tradition was Tomasello's (1992) 'Verb-Island Hypothesis' (VIH). This was based on a diary study of his English-speaking child, T, between the ages of 15 and 24 months. T's early language use was conservative, with little overlap in the way she used different verbs but much continuity of use within a single verb. There was no evidence of a verb general category. Some verbs were never marked for tense (past tense morphology in English) or aspect (the use of auxiliaries and morphology to indicate the nature of a situation, e.g. whether it is complete or ongoing, fixed or changing, temporary or of long duration), others were marked only for one or only for the other, and very few (2%) were marked for both functions. An example is that T first used *spill* in its past tense form only but *drop* in the present-tense only, before they later became used in both ways). Tomasello's (1992) VIH was that children learn language conservatively and initially acquire individual verbs, tied to the structure in which they appear in the input. Very young children (2 years) lack any abstract knowledge of either verb categories or constructions. Their language use is built around verb-specific schemas with open nominal slots, for example young children would represent the transitive verb '*hit*' as:

Noun Phrase (hitter) – form of verb HIT – Noun Phrase (hittee)

The adult-like verb-general entities 'subject', 'object' and/or 'agent' and/or 'patient' are replaced by the verb specific 'hitter' and 'hittee', 'pusher' and 'pushee' and so on. Children acquire individual verb meanings incrementally and only later do they begin to form abstractions by generalizing across verb forms.

Lieven, Pine and colleagues (e.g. Pine and Lieven, 1993; Pine, Lieven and Rowland, 1998; Theakston, Lieven, Pine and Rowland, 2001) later showed that the VIH's focus on verbs was too narrow. Pine, Lieven and Rowland (1998), for example, investigated the extent of lexical overlap shared between different instances of grammatical construction types in the first 6 months of twelve English-speaking children's multiword speech. They found that although most of the children's language was grammatical, much could be explained by a relatively small amount of lexically specific knowledge. On the one hand, Tomasello's research was supported in that the children exhibited a lack of overlap in the verb types to which they applied different morphological markers, lack of overlap in the verb types with which they used different auxiliaries, dominant usage of the first person singular nominative pronoun (I), and lack of overlap in the lexical items used as subjects and direct objects of transitive verbs. On the other, children did show some grouping of verbs into relatively narrow subgroupings based on low-scope frames with slots, for instance "*I'm VERB-ing it*". However, this lexically-specific use of grammatical items provides no evidence of truly abstract syntactic categories or schemas. Rather, the data suggest that young children's knowledge is non-general and organized not only around verbs but also around other high-frequency markers, especially pronouns.

More recently Lieven, Behrens, Speares and Tomasello (2003) were able to explore in further detail the form children's early lexically organised schemas may take by analysing a dense dataset from a single 2-year-old fluent girl, Annie . They assessed the relation between one sample of Annie's creative utterances (one hour of interaction with her mother at age 2;1.11) and those that she had previously produced (the previous 6 weeks of samples at a rate of 5 hourly sessions per week, and the accompanying maternal diary of novel utterances). In the hour-long sample at age 2;11.1. there were 295 multi-word utterances of which 37% were 'novel' (not previously produced in their entirety). These novel utterances were compared with utterances from the previous six weeks of the same child's

data using a 'morpheme-matching' method to identify how each novel utterance in the final session differed from its closest match in the preceding corpus. This systematic method was used to identify schemas in Annie's speech, by considering the number and order of the morphemes that were shared between utterances and the frequency of utterances of a similar form. If a number of utterances was found with the same positive morpheme match and the same overall number of morphemes, but with type variation in the same position as the target utterance, then this was defined as a schema, for example 'I got the butter' and 'I got the door' (matching morphemes underlined) would constitute instances of a 'I got the W' schema (W indicates a word that varies across utterances).

Lieven et al.'s (2003) data suggest that the apparent high degree of creativity exhibited in the naturalistic speech of young English-speaking children could be at least partially based on entrenched schemas and a small number of simple modifying operations. There was a very close relationship between Annie's novel utterances and her preceding utterances, with most requiring only the substitution or addition of a single word. These findings do not seem to be accounted for by the repetitive nature of child-parent conversation, because the application of the same process to Annie's mother's speech revealed that though she produced a number of novel utterances in the last session comparable with Annie's, the preceding corpus did not provide as many prior schemas and exemplars, and more of the matches required multiple and more complex operations, such as insertions and rearrangements. Lieven et al.'s (2003) analysis can be directly related to children's on-line language production (see discussion in section 3 below).

In summary research using children's naturalistic data is highly suggestive of a period in early language development when children are building up low-scope patterns on the basis of their pragmatic understanding, what they want to say and relative frequencies in the input. For sure, children are creative with language from the beginning but the UB approach suggests that the scope of their creativity changes from initially low-scope, item-based representations and slowly develops into the more general and abstract knowledge represented by adult language. However, to establish this experimental research is required that can control what it is the children are hearing and identify how their generalisations are being made.

2.2. Experimental Data

Naturalistic data are needed to provide insight into the language patterns that children actually choose to use in their speech, which is particularly useful for studying which factors affect fluency. They provide language in context rather than artificially isolating words or structures which could miss certain features of the data by narrowing the field of observation too much (e.g. ignoring relevant constructions) or making the stimuli too hard or too easy (e.g. the child who stutters either exhibits no disfluencies or cannot say the words at all). This said, it is crucial to back up naturalistic findings with experimental evidence when the concern is to identify how abstract children's early representations are. Experimental work is the only way to test predictions and identify causal links. Any conclusions from naturalistic data are correlational and carry the risk of suggesting spurious causal relations created by hidden variables. Several inventive experimental methodologies have been developed. Each reveals different aspects of children's knowledge and, in addition, different methods work better for different ages. In most paradigms, the data can only indicate how abstract children's representations are if nonce material is used to systematically control the input that children receive. This allows the language variables that affect performance to be teased out and determine how the language input affects the language output. In contrast, using familiar words is subject to the same problems as naturalistic data (i.e. the words are likely to have been heard by the child already). Only by using novel nonsense (nonce) words (e.g. '*dacking*') can the investigator control the ways in which the child has already heard the words used and be sure that any productive use (i.e. in a construction different from that which they have heard in the experiment) stems from at least partially abstract generative patterns.

One experimental paradigm used to test comprehension is 'acting out'. Experiments have shown that from early in development young children show the ability to act out some types of sentence appropriately, for example, the English transitive, but only when they use familiar verbs, not unknown verbs (e.g. Roberts, 1983). When 2-year-olds are taught a novel action paired with a novel verb (e.g. '*This is dacking*') and are then asked to '*Make X dack Y*' or '*Show me: X is dacking Y*' they are equally likely to make either X or Y the agent of the action (Akhtar & Tomasello, 1997).

Acting-out can only be done with children from about 2 years of age. Before this the preferential looking methodology is often used. This uses infants' looking patterns to infer their ability to distinguish utterances with contrasting syntax. Using designs of this kind, researchers have found that some 2-year-olds seem to be responsive to some aspects of transitive constructions in English if they are used with verbs that they know. For example, Naigles (1990) showed that when children as young as 2;1 heard transitive utterances with known verbs they preferred to look at one participant doing something to another (indicating a causative meaning) rather than two participants carrying out

synchronous independent activities (intransitive meaning). It is important to note, however, that this study does not tell us that 2-year-olds possess a full awareness of the functional role of the verb (for example to be able to connect the pre-verbal position with the subject and the post-verbal position with the object) because the test was conducted with familiar verbs. The children could simply be working on the basis of the patterns in which they had previously heard the verb used rather than working on a verb-general template of the transitive. Only Fisher (2000) has used unknown verbs in the preferential looking paradigm, using sentences like "*The duck is gorping the bunny up and down*". However, the sentences she presented to children (1;9 and 2;2) contained prepositional phrases that provided additional information that children could use to interpret the meaning of the sentence instead of using the formal syntactic marking (see also Fisher, 1996, 2002). The precise nature of the knowledge that children use in these preferential looking tasks and how abstract it may or may not be is still under intense investigation.

There is also a large and growing body of evidence on children's early linguistic knowledge from production experiments using novel verbs. Most of these focus on children aged between 2 to 3 years because of the significant developments in grammar that occur at this time. The overall finding is that 2-year-old children's early productivity with syntactic constructions is highly limited. This confirmation comes from studies like that of Tomasello and Brooks (1998), who exposed 2- to 3-year-old children to a novel verb used to refer to a highly transitive and novel action in which an agent was doing something to a patient. In the key condition the novel verb was used in an intransitive sentence frame such as "*The sock is tamming*" (e.g. a bear was doing something that caused a sock to 'tam', which was akin to *rolling* or *spinning*). Then, with novel characters performing the target action, the adult asked children the question "*What is the doggie doing?*" (the dog was causing a new character to tam). Agent questions of this type encourage a transitive reply such as "*He's tamming the car*", which would be creative since the child had only heard this verb in an intransitive sentence frame. The results showed that very few children produced a full transitive utterance with the novel verb. As a control, children also heard another novel verb introduced in a transitive sentence frame, and in this case virtually all of them produced a full transitive utterance, demonstrating that they can use novel verbs in the transitive construction when they have heard them used in that way. Moreover, 4- to 5-year-old children are quite good at using novel verbs in transitive utterances creatively, demonstrating that once they have indeed acquired more abstract linguistic skills children are perfectly competent in these tasks (Pinker, Lebaux & Frost, 1987; Maratsos, Gudeman, Gerard-Ngo & DeHart, 1987; see Tomasello, 2000, for a review).

In another novel verb experiment with young children (aged 2;5 to 4;5), Akhtar (1999) used the novel verb to describe a transitive action but used it with a novel word order, for example "*The bird the bus meeked*". When the younger children were given new toys and encouraged to talk about the new event, they quite often repeated the pattern and said things like "*The bear the cow meeked*". On the other hand, the 4-year-olds consistently corrected the pattern they had heard to canonical English word order in their responses (e.g. correcting to "*The bear meeked the cow*"). These findings are consistent with the hypothesis that when 2- to 3-year-olds learn about *meeking* they only learn about the order of arguments for *meeking* and do not assimilate the newly learned verb to a more abstract, verb-general linguistic category or construction that underpins the canonical English transitive. However it is clear that this is a developmental process in which knowledge of the transitive slowly builds up. Thus the 2-year-olds were better at repeating sentences with novel verbs if they had heard these verbs used in SVO order than if they had heard them used in the ungrammatical (for English) orders SOV or VSO. This suggests some sensitivity to conventional usage. Abbot-Smith, Lieven and Tomasello (2001) obtained similar results for younger children with the intransitive construction. When children aged 2;4 were presented with a verb they knew (e.g. *jump*) in noncanonical word order they corrected the word order they heard to canonical word order even more than the 2-year-olds in Akhtar's study, but they were more likely to use the ungrammatical word orders that they had heard with novel verbs. However, they too more readily reproduced canonical orders than noncanonical orders. This suggests that they possess some knowledge of an abstract transitive construction but that this is not yet strong enough to use to override the patterns they receive in the input.

Overall, the results from the different methodologies indicate that between 2 and 3 years of age, young English-speaking children are still in the process of building up their abstract, verb-general constructions. Similar evidence also exists for other languages (e.g. Allen, 1996, for Inuktitut; Pizutto and Caselli, 1994, for Italian; Rubino and Pine, 1998, for Brazilian Portuguese). The results so far mainly relate to what children lack before 3 years of age with little implication about the ways in which they are productive with language in their multi-word utterances. The naturalistic data discussed above (Tomasello, 1992; Pine & Lieven, 1993; Pine, Lieven & Rowland, 1998; Theakston, Lieven, Pine & Rowland, 2001; Lieven, Behrens, Speares & Tomasello, 2003) suggest that though 1- and 2-year-olds lack abstract syntactic knowledge in the sense of UG competence, they are able to use a more rudimentary grammatical system of lexically-based productive patterns, like '*More X*' or '*I'm Xing it*'.

Experimental evidence exists to back this up. Data-driven learning theorists reason that these 'slot-and-frame' schemas could be derived from a combination of the repetition and systematic variation of phrases that have been found to occur in the input (Cameron-Faulkner, Lieven & Tomasello 2003). The consistent parts could be imitatively learned whilst the more abstract slots could be formed on the basis of the communicative parallels across varied instances. The *X* in '*I'm Xing it*', for example, refers to a different action in each instance but its communicative function is constrained by that of the whole schema and always refers to acting on an object.

Childers and Tomasello (2001) investigated the building of slot-and-frame schemas experimentally. They trained 50 children aged 2;6 with several hundred transitive utterances using either 16 familiar or 16 unfamiliar English verbs (as measured by whether or not they appeared for 2-year-olds on the MacArthur CDI; Fensen, et al., 1994) spread over 4 sessions in as many days. For some children, the agent and patient in all sentences were labelled with only nouns and for other children they were labelled with both nouns and pronouns. The children then saw 4 novel actions and heard 4 novel verbs in non-transitive constructions (2 intransitives and 2 passives, using one N and one PN in each). To see if they could produce a transitive with the novel verb, children saw each novel action modelled again and were asked both neutral and transitive-pulling elicitation questions.

The authors found that, regardless of whether the verbs used during training were familiar or unfamiliar, children were best at generalising the transitive construction to the novel verb if they had heard both pronouns and nouns during training rather than nouns only, for example they generalized better after hearing "*Look! The bear's striking the tree. See? He's striking it*" than after hearing "*Look! The dog's hurling the chair. See? The dog's hurling the chair*." This suggests that the children either learned a more abstract transitive schema from which to generalise to new verbs by hearing variation in the subject and object slots, or that they learned a pronoun-based transitive schema. One feature that supports the latter conclusion is the fact that after the pronoun training condition, children mostly used pronouns in their responses. This supports the suggestion raised above when discussing naturalistic data that pronouns are important in forming schemas.

Further experimental evidence that children's syntactic knowledge about transitives is arranged around pronouns comes from a priming paradigm (Savage, Lieven, Theakston and Tomasello, 2003; submitted). Savage and colleagues recently adapted the 'syntactic priming' paradigm and used it for the first time with young children. The paradigm is well established in the adult literature (e.g. Bock 1986; Pickering and Branigan, 1999; Schenkein, 1980). It refers to the phenomenon whereby processing an utterance with a particular syntactic structure facilitates the processing of a subsequent utterance with the same or a related syntactic form, even though alternative forms would be semantically appropriate. An example would be when a speaker chooses the passive over the active after producing another passive recently. The consensus view is that that syntactic priming reflects underlying syntactic knowledge and as such it provides a methodology to tap into speakers' underlying knowledge (e.g. Chang, Dell, Bock and Griffin, 2000; Pickering, Branigan, Cleland and Stewart, 2000).

For adults, a stronger priming effect is produced when the primes and targets share the same verb than when they do not (Pickering & Branigan, 1998). Having said this, the effect is significant even when the verb differs between the primes and targets (Pickering & Branigan, 1998). The latter finding, in particular, demonstrates that priming is effective over and above an effect of lexical overlap and indicates that adults possess abstract knowledge of the primed construction. For children, however, the pattern is different. Young children (3-4 years) only show a priming effect when there is much lexical overlap between the primes and targets, for example by priming them with a slot-and-frame type schema like '*It got V-ed by it*' for the passive, where they need only slot the target verb into the lexical pattern to be primed. Later on however (by around 6 years), children are primed in a more adult-like fashion. They show an effect both in the high lexical overlap condition and also at a more abstract level, for example '*The bricks got pushed by the digger*' will prime '*The cake got cut by the knife*', where both the N and V slots vary between prime and target. These data confirm that young children lack adult-like abstract syntactic knowledge, but interestingly they also indicate that they do possess some verb general knowledge. They are thus in line with the naturalistic data of Pine, Lieven and colleagues that children's early syntactic knowledge takes the form of slot-and-frame schemas that are anchored by lexical items like pronouns.

To conclude this section, the experimental evidence converges with the naturalistic data in indicating that children's early syntactic knowledge takes the form of slot-and-frame schemas. Only in the months leading up to 3 years do children start to form relations between these schemas and build more abstract knowledge of constructions. The evidence suggests that children of 3 years are reorganising grammar from lexically specific schemas into more abstract ones. This coincides with the age at which children become disfluent. This supports the suggestion made earlier that stuttering could therefore be a 'side effect' of the increased demands that this reorganization places on the child's processing system. In the following sections, the changing nature of children's syntactic knowledge across development is

considered and how this may relate to their fluency behaviour in terms of the online speech mechanisms involved.

3. Low-Scope Schemas and how they Relate to Fluency in Speech Production

The relationship between linguistic representations and language production in real time is complex. A growing body of evidence exists that suggests that adults possess more than one route of access to their underlying representations during speech (e.g. Bybee & Scheibman, 1999), meaning that as well as being able to generate novel utterances from abstract representations they also seem to produce 'frozen' strings of language more directly without using abstract grammar, so called 'unanalysed' units. One example is 'grammaticalization', the diachronic process by which frequent usage allows conventionalized structural expressions to emerge in a linguistic community (Bybee & Hopper, 2001). Over time the link between a phrase and its underlying form is lost, for example '(be) supposed to' has lost its passive status. In its reduced form the infinitive is phonologically fused to the verb such that a passive agent seems grammatically unacceptable, as in 'He's s'posed to be very knowledgeable *by most people' (where asterix means grammatically unacceptable) (Bybee & Thompson, 1998, p. 2). The position this issue is given in theories of language production depends on the perspective taken, but its importance is recognized in both UB and generativist theories. Generativist theories retain the central place of abstract representations but recent versions do not simply ignore direct access as being peripheral. Some authors have become increasingly involved with explaining the relationship between generativist grammar and the role of the 'exceptions' (idioms, partial idioms and low-scope constructions) (e.g. Lebeaux, 1988; Jackendoff, 1996; Culicover, 1999). However, in UB theories of grammar, the idea of different degrees of abstraction is an integral component, as will become clear below.

The UB approach has its roots in cognitive and functional linguistic theories that view grammar as a response to discourse needs. As a result grammar is dynamic and experience-driven (e.g. Bybee, 1998; Goldberg, 1995, 1999; Hopper, 1987; Langacker, 1987). These authors characterise grammar as an inventory of language specific 'constructions', or schemas, for example the 'get' passive in English could be 'NP got V(past tense) by NP'. Some authors would argue that these constructions can be at a highly abstract level that parallels the abstract nature of traditional grammar (though, importantly, they are not identical because they are still language-internal rather than universal) (e.g. Bybee, 1998; Ono and Thompson, 1995). Others maintain that a high level of abstraction is unlikely (Croft, 2001). Ultimately these two positions will have to be resolved empirically. The important point is that UB theories all agree that redundancy exists in underlying representations. That is, the existence of a directly accessible, construction-level representation (multiple words stored as one unit) does not preclude the co-existence of smaller level units (i.e. words) that can be used to generate sentences when a more unusual linguistic composition is required. Language users might be able to coincidentally recognise both construction level units and their individual component parts, and the construction and its units may be interlinked in memory by means of a 'network' of representations (Bybee, 1998).

The most unanalysed chunk at the most concrete level of representation for adults would be an idiom, such as '*He kicked the bucke*t' to mean '*He died*' (and possibly some additional semantic nuances), which cannot be understood by combining the individual words but can be recognized as containing analyzable components that appear elsewhere in similar syntactic positions. Note that while this is an idiom it can be used in interaction with other constructions to produce novel utterances, for instance in different tenses (He has kicked the bucket) or utterance-level constructions (Did he kick the bucket?) Another example is the various levels at which a person could represent the passive. The 'get' passive could be represented as a concrete, utterance-level representation without a by-phrase (a set phrase derived from the input, such as '*It got broken*'). It could also occur as a partially abstract item, with some filled (in bold) and some open slots as in '**It got** Verb-**ed**'. This representation could in turn be connected in memory to that of the by-phrase so it could occur both with and without it (It got verb-ed by X). At the highest levels of abstraction the construction would be represented at a more general level for all component parts and no concrete parts, and it would be connected to the be-passive or even to other constructions (Tomasello, 2003), for example 'N be/get (any tense) Ved (by N) (*He was eaten by a dragon; He got run over by a bus*). From the 'network' perspective (e.g. Bybee, 1998), grammar exists as an interlinked inventory of constructions ranging from fully concrete, to partially concrete, to fully abstract. The links between concrete phrases represent the building up of levels of abstraction, but the sub-units can also be accessed directly, depending on how creative and original an utterance the speaker wishes or needs to generate. On occasion, it may be that there is nothing directly useful in the existing inventory, or that it cannot be recalled. This might be how errors of commission and overextension arise. The network would function via a sort of 'path of least resistance' operation for language production.

When adults or children produce speech in real time (online), a complex interaction takes place between the different routes of access to representations. The relevance of the UB models to fluency research is that they should be able to generate concrete predictions about where disfluency is most likely to occur (i.e. in which parts of a construction and in which forms of construction). Which phrases have to be built up and which can be accessed directly would impinge on fluency. In particular, the more concrete phrases will require less effort from the language production system (making them less prone to disfluency) than the more creative utterances. The more frequently used and more concrete phrases demand only a relatively direct and automated retrieval step, whereas more creative productions require that certain abstract slots are filled with lexical items. The moments in speech that precede the more abstract parts of an utterance will be more vulnerable to disfluency as a consequence of the greater demands placed on the language production system at that point, in preparation for the subsequent part of the utterance. The loci of these vulnerable moments will vary for adults and children, because children's representations are gradually changing over time.

The findings from Lieven et al. (2003) that were discussed above suggest that in children's on-line language production, concrete strings stored in memory might interact with more abstract categories that have traditionally been the central focus of linguistic study. The knowledge required to underpin Annie's language processing could be a combination of strings stored in memory that provide lexically-specific 'frames' and categorical knowledge that is used to fill 'slots'. Lieven et al. (2003) tentatively propose that the types of schemas they identify are psychologically realistic in terms of how a child constructs novel utterances on-line, constituting storage and planning units. Likewise, they propose that the operations they identified may represent psycholinguistic operations that children use to manipulate their representations and construct novel utterances on-line. Examples are 'Substitute' (using the same lexical pattern with one item substituted for another, as in the formation of "*I have some toast*" from "*I have some coke*") and 'Add-on' (the same lexical pattern plus one added item, as in the formation of "*Put a bit more here*" from "Put X" and "*A bit more here*")[1]. Whether these operations exist is an empirical question. One way of exploring it would be to investigate the impact of the operations on speech fluency. This would involve using data on stuttering to inform theories of fluency development. Conversely, though the schemas and operations identified by Lieven et al. (2003) were found for a fluent speaker, their analysis could provide an interesting new angle on how to explain fluency and disfluency patterns in speech production also for children who stutter. A sophisticated and empirically-based model of speech production in children could be developed by considering how the schematic operations suggested by Lieven and colleagues interact with the child's developing mastery of phonetic complexity (e.g. MacNeilage & Davis, 1990; Davis & MacNeilage, 1995; Jakielski, 1998) and phonological representations.

The above proposals are as yet unsubstantiated. The empirical evidence to date for UB models has focused mainly on the nature of the underlying linguistic representations that children and adults possess and not how they use these to produce speech. It is still an open question as to whether the models of representation have any real psychological significance in online speech production. There are, however, a few studies that relate the UB approach to online processing in adults. Schilperoord and Verhagen (1998) studied the timing of speech production when adults read a passage of text. They showed that the locations at which their subjects segmented the text, indicated by pausing, did not always coincide with traditional grammatical boundaries, for example in the case of restrictive relative clauses, and subject and complement clauses. Schilperoord and Verhagen argued instead that the way in which the speaker makes conceptual links between clauses drives their segmentation of the text (see also Verhagen, 2001). For the restrictive relative clause, "if a constituent of a matrix-clause A is conceptually dependent on the contents of a subordinate clause B, then B is not a separate discourse segment" (Schilperoord & Verhagen 1998, p. 150) (for an example see 3 and 4 below). This echoes Langacker's (1991) observation that for restrictive clauses the speaker can only conceptualise the referent in the matrix structure once they know the contents of the relative clause. This means that whether or not the matrix clause can be considered to be an independent unit of processing depends on the nature of the subordinate clause. To illustrate this, compare the texts in examples 3 and 4, below:

3. These schools all appear to have relatively many students who grew up in culturally deprived families
4. They shouted at the waiter, who so far did not seem to have noticed them[2]

For the restrictive clause in (3) we depend on the contents of the relative clause ("who…") to understand the referent of 'students', because it provides necessary information. For the non-restrictive

[1] Examples taken from Lieven et al.'s (2003) analysis of Annie's data.
[2] Examples 3 and 4 both taken from Verhagen, 2001

clause in (4) we can conceptualise the referent of 'the waiter' independently of the relative clause, because the relative clause simply provides extra information.

Schilperoord and Verhagen's model provides a usage-based alternative to formal grammatical segmentation in speech because the linking of clauses is based on what they actually mean to the speaker rather than on abstract grammatical rules that are independent of lexicon and semantics. Moreover, the evidence suggests that the conceptual boundaries constitute a psychologically real unit of speech processing that traditional grammatical boundaries do not recognize. Segmenting a text according to conceptual conditions predicts the segmentation of text in spoken Dutch as measured by the good correlation between segments and pausing patterns (Schilperoord, 1996, 1997).

Further evidence that is consistent with Schilperoord and Verhagen's theory is provided by Gee and Grosjean (1983). These authors did not look directly at functionally derived units; their units of analysis were 'prosodic bundles' that were derived from an algorithm based on formal linguistic principles, but importantly these units did not correlate with syntactic boundaries. In fact they closely approximated the 'phonological word' unit used by Howell and colleagues (discussed in section 5) that is made up of a content word surrounded by function words. Gee and Grosjean found that that pausing precedes a function-content word unit but hardly occurs at all at the boundary between the function and content words. The study can be interpreted, then, as showing that speakers pause before they embark on one of these function-content word units, and doing so avoids the risk of repetition of function words. Related to the evidence from Gee and Grosjean's (1983) work, Pinker (1995) gave examples showing that pauses do not always occur at a major (formal) syntactic boundary. In the examples used to illustrate this, the pauses do, however, all occur at phonological word boundaries. So, again, this is consistent with speakers pausing prior to a planning unit that is not derived from formal syntax.

The naturalistic data are backed up by some experimental evidence. Vogel Sosa and MacFarlane (2002) used a word-monitoring paradigm to compare the reaction times of adult English speakers for recognizing the function word '*of*' in frequent versus infrequent collocational contexts. They found that speakers responded slower to '*of*' when it occurred in a highly frequent collocation, such as '*kind of*' compared to a lower frequency collocation like '*sense of*'. This is consistent with the UB theory that words that very often occur together in connected speech are stored as a single unit and accessed holistically, as suggested by the UB model. To recognise '*of*' in a frequent collocation would require decomposing the holistic unit into its individual parts, similar to the morphological decomposition required to access the past tense morpheme after hearing '*walked*'. The words in the holistic unit would only be recognized via their connections with other instances of the component parts '*kind*' and '*of*'. The less frequent collocations would not be fused into units and so the recognition of '*of*' would be direct, and hence faster. Though the data from Vogel Sosa and MacFarlane focus more on lexical access than syntactic processing, they do provide some preliminary evidence that UB units are relevant in real time speech processing. They suggest that speakers do not plan their speech in segments based on the units of formal linguistics, but rather plan in segments that derive from the input.

To summarize, UB theories predict that we can expect words to be stuttered at processing boundaries and these may not be the traditional grammatical boundaries of formal linguistics. Notably, the boundaries will vary over developmental time as children build up more abstract representations. The problem is that because the UB approach is still young, it has tended to focus mainly on the nature of the underlying grammatical representations and little is yet known about how the different routes to production (direct and via abstract units) interact online. Though the evidence discussed above is consistent with the proposal, there is little evidence that directly assesses what the alternative units of processing may be. The neglect is most noticeable in the child language literature, where no real theory has yet been provided from the UB perspective to explain the online mechanisms involved in children's speech production. In other areas of study, however, the interest in early language production and fluency has been strong. There has been much debate about what factors affect the location of disfluencies in speech (e.g. grammatical complexity) and the specific nature of the mechanisms of language production that are entailed in this. We will turn now to a discussion of the existing literature surrounding these topics, before returning to the UB approach to explore how the two fields may be able to contribute to one another.

4. The Wider Relationship between Linguistic Representations and Fluency in Speech Production

The relationship between natural language factors and stuttering is widely researched in children who stutter (Wingate, 1988; 2002). Generally speaking (and on first sight somewhat counter-intuitively), fluency problems are evident on function words (Bloodstein & Gantwerk, 1967; Bloodstein & Grossman, 1981) and involve repetition of the whole of these words (Conture, 1990), e.g. '*at at at school'*. Young speakers seem not to be influenced by the phonetic structure of the words (e.g. whether

the word contains a consonant string or not, or whether the consonants in a word have manners that are difficult for a child to produce such as fricatives and laterals) (Howell, Au-Yeung & Sackin, 2000). In contrast, fluency problems in adults who stutter are evident on content words (Howell, Au-Yeung & Sackin, 1999) and the disfluency often involves the first part of these words (Conture, 1990), e.g. '*at ssssssschool'*. Their fluency is affected when the content word has properties that make words difficult to acquire for children, e.g. '*school'* would be difficult because it has a consonant string (sk) containing a late emerging consonant (s) in word-initial position (Howell, Au-Yeung & Sackin, 1999).

Different approaches have been taken to explaining why childhood stuttering is anomalous. Wingate (2002) argued that the childhood pattern is just normal nonfluency, not stuttering (which would also explain why there is a high rate of recovery from childhood 'stuttering'). The majority of authorities maintain that there is developmental change (Conture, 1990; Howell, 2004). Howell goes further and argues that the different patterns of stuttering reflect contextual influences. Word repetitions are stuttering-like disfluencies that precede difficult words. Howell argues that the childhood form of disfluency is associated with getting the words ready in time (see later for further details).

The empirical evidence concerning the relationship between syntax and fluency remains equivocal. The joint questions of whether fluent children (children who do not stutter: CWNS) use more complex syntactic structures than children who stutter (CWS) and whether CWS have deficient syntactic capacity have yielded contradictory findings. Most authors agree that children are more likely to produce disfluencies on utterances that are more syntactically complex (see Bloodstein, 1995; Karniol, 1995; and Ratner, 1997, for reviews) but there is wide variation as to how this relationship should be interpreted. Many authors hold that stuttering is the external symptom of the difficulty that CWS experience with syntactic processing, and that this is evidenced by a correlation between syntactic complexity and disfluency rate (e.g. Blood & Hood, 1978; Bernstein, 1981; Brutten & Hedge, 1984; Gordon & Luper, 1989; Gaines, Runyan & Meyers, 1991; and see Karniol, 1995). Others argue that the interpretation of these conclusions is a less straightforward matter. Yaruss (1999) also found that disfluency was related to syntactic complexity, but showed through logistic regression that this was a poorer predictor of stuttering than was length of utterance, and that neither of these measures was a particularly strong predictor. He argued that these factors cannot adequately account for stuttering by themselves. Logan and Conture (1997) argue that the relationship they found between stuttering and a higher number of clausal constituents in the utterances of CWS (regardless of length of utterance) does not necessarily reflect syntactic processes, but could instead reflect prosodic planning. An argument has been made that CWS are less syntactically able than CWNS. Some studies show that CWS use less complex utterances than CWNS (e.g. Wall, 1980; Wall, Starkweather & Cairns, 1981; Howell & Au-Yeung, 1995) but others have found no difference between CWS and CWNS on measures like the Developmental Sentence Analysis (Westby, 1974) and some even suggest a relative level of syntactic precocity in CWS when compared to norms (e.g. Ratner & Sih, 1987; Watkins & Yairi, 1997). The relationship is indeed complex. Westby (1974) found that CWS make more grammatical errors than CWNS but Yaruss (1999) found no difference in grammatical accuracy of the stuttered and fluent utterances of CWS.

Some of the confusion regarding the relationship between syntax and fluency results from the wide variation in how different studies appraise syntactic complexity. Several different measures have been used in various combinations by investigators to classify children's productions in naturalistic data. They include the number of clausal constituents a sentence contains (e.g. Logan & Conture, 1997), the number of clauses in an utterance (Yaruss, 1999; Wall, 1980), the amount of embedding in an utterance (Kadi-Hanifi & Howell, 1992) and the earliest age at which different structures are produced by children (e.g. passives and negatives occur later than actives and declaratives, so are considered harder) (e.g. Silverman & Bernstein-Ratner, 1997; Yaruss, 1999). Methods of assessing children's underlying knowledge have also been used, including the Reception of Syntax Test (ROST) (Howell, Davis & Au-Yeung, 2003) and elicited imitation paradigms wherein children try to reproduce sentences that comprise various levels of syntactic complexity (Silverman & Bernstein-Ratner, 1997).

The methodological disparity inherent in the literature can partly explain the confusing pattern of results but there is, in fact, a hidden assumption that is shared by all of the methods mentioned. They all characterize syntax in the traditional sense, with all exemplars of a general syntactic category being treated with equivalence (a possible exception being Kadi-Hanifi & Howell, 1992, who also drew on semantic factors). None of the methods examined systematically the relationship between the syntactic frame that is used by a child and the lexical content therein. The importance of this should be clear, especially for young children (2-3 years), from section 2 where we considered the nature of children's early syntactic knowledge as revealed by the UB approach. For example, if passives are classified as more complex than the actives because they emerge at a later stage of development in English, their lexical composition still cannot be ignored. If a child produces a passive utterance, we have already seen that we cannot be sure how they have gone about doing so. A passive could be constructed using a

single operation to modify a highly concrete schema like *'The W got broken'* (where W means a slot that contains a variable word). This would actually be easier to construct than a more creatively produced active construction, as when a child inserts words and rearranges them to form *'Mummy's pushing me now'* from *'Mummy's trying to push me'*. This may be the case even when utterance length is the same. As shown, a classification of children's utterances based on adult-like abstract categories does not necessarily represent accurately the capacities of young children. Instead, a more appropriate and representational measure would be how creative the utterance is, for instance how abstract the component parts are (e.g. is the construction used with only one verb or many different ones?) and the number of manipulations needed to reach the present utterance from previous ones (i.e. compare one structure with all preceding instances).

A related and crucial issue is the inability of absolute measures to take account of the changing nature of children's syntactic knowledge across age. For example a measure that classifies all passives as complex will attain different results for a child as they grow older, because early passives are likely to actually be more simply produced from lexical schemas, whereas later instances are likely to be produced from more abstract underlying constructions. The UB approach has assembled a large body of evidence about the way in which children's syntactic knowledge changes as they grow older and this knowledge is ripe for application to fluency research. To exploit this potential, we need first to find a suitable model of the processes involved in speech production in real time. This will allow us to link together the UB approach insights into the units of linguistic representation with the processing mechanisms that act on these to transform them into the spoken speech signal.

5. The EXPLAN Model of Speech Production (Howell & Au-Yeung, 2002)

The EXPLAN speech production model of Howell and Au-Yeung (2002) is unique in being the only model to be explicitly developmental, and to explain the high incidence of the anomalous stuttering on function words in early development. EXPLAN arose out of the work of Howell, Au-Yeung and Sackin (1999) who investigated the relation between stuttering on function and content words within a contextual unit, the phonological word, which specifies the extent of units that incorporate the two types of word (see also Au-Yeung, Howell & Pilgrim, 1998, who first introduced phonological words into analysis of stuttered speech).

Phonological words (PWs) as defined by Au-Yeung et al. (1998) have an obligatory content word and a variable number, from zero up, of function words preceding and following it. An example is *'I split it'* which has one function word before, and one after, the content word (the verb "split"). The function words in a PW are associated with their content word by sense unit rules (i.e. the function word has to be semantically related to its content word). Three properties of stuttering are seen in PWs that have function words before and after the content word. First, stuttering on function words, on the vast majority of occasions, occurs on those that precede the content word (on *'I'* in the preceding example) (Au-Yeung et al., 1998). Second, stuttering occurs either on the function word or words that precede the content word or the content word itself, not both (Howell et al., 1999) – you see *'I, I, I split it'* or *'I sssplit it'* commonly but rarely see *'I, I, I sssplit it'*. Third, the tendency to stutter more on content words as age increases is associated with a corresponding decrease in stuttering on initial function words (Howell et al., 1999). A fourth important feature, not specifically about the distribution of stuttering, is that the type of stuttering on function words tends to involve hesitation around, or repetition of, the whole function word whereas stuttering on content words typically involves difficulties producing the first part of these words as in part-word repetitions (*'s..s.. split'*) or prolongations (*'sssplit'*) (Conture 1990). Another relevant finding is the work by Gee and Grosjean (1983) mentioned above, showing that pausing precedes a function-content word unit, but hardly at all at the boundary between the function and content words.

The EXPLAN model offers an account of all of these findings on the assumption that at the root of all stuttered output is the difficult word (that for English is usually a content word). In the EXPLAN model, fluency control problems arise because the complex content words can take too long to generate for the context in which they need to be produced. In a connected stretch of speech, like the PW in the earlier example, the plan for the content word may not be ready for execution immediately after the initial function word has been completed and, therefore, the whole content word cannot be executed. The speaker can do one of two things to deal with this problem: First, the speaker can interrupt speech by pausing or re-executing the words that precede the word that is not ready (i.e. the initial function word or words). Pausing or repeating gains time to complete the plan of the content word, but they will only work when these disfluencies occur on the initial function word in an example like *'I split it'* (accounting for the first feature noted above). To be concrete, *'I split it it'* occurs rarely because repetition of *'it'* cannot gain more time for planning a content word that precedes it (*'split'* in this case). When word repetition or hesitation around the initial function words occurs, it provides more time that, in turn, prevents stuttering on the content word. This explains why in any particular disfluent PW,

stuttering happens in an either-or fashion on initial function or content words (the second feature noted earlier). Finally, the repeated function words are produced in their entirety as their plan is complete.

The second possibility for the speaker is to start the utterance after the initial function word has been produced, even though its plan is not complete. The part of the plan that would be available is the initial part (assuming speech is generated left to right). There is then no function word repetition or hesitation, rather stuttering occurs on the content word. If the plan runs out (feature two), the first part is all that can be produced as that is all that is available (feature four). To account for the developmental changes, Howell and Au-Yeung (2002) assumed that as speakers get older, they change from responding to situations where the plan cannot be generated in time using function word repetition to producing parts of content words (feature three). The issue as to what underlies this change has not been worked on to date (though see Howell, 2004, for some hypotheses).

6. The Relationship between EXPLAN and the UB Approach

EXPLAN provides us with a developmentally orientated model of online speech production that takes account of the planning context of the stuttered units. However, the PW as the unit of speech production is not derived empirically but from phonological theory (Selkirk, 1984). Naturalistic evidence that adopts it as a unit of analysis fits well with the EXPLAN model but is this because the unit is psychologically real or does it simply correlate with other units that are, namely those identified by the UB approach? Some researchers would argue that the difference between the UB units and the PW resides in a distinction that can be drawn between 'linguistic' and 'speech output' representations. This is a matter of theoretical perspective but at least in the case of children there is no empirical evidence to show that this distinction is necessary or justified.

The PW is an abstract component of generative phonological theory in the same way that traditional clausal boundaries and constituents are abstract components of generative syntactic theory. In fact, the need for generative phonology only arose because generative syntax failed to explain why the phonological and prosodic groupings of spoken language do not always match up with syntactic boundaries. Formal linguistics developed phonological theory as a further layer in the language production process to account for these patterns (Jackendoff, 2002). In fact, viewed as a whole, the generative model is rather clumsy and, in its attempt to fit theory to data by addition rather than modification, it loses the elegance that its early supporters admired. UB models are more parsimonious and maintain that syntax is not completely isolated and is connected to semantic and phonological knowledge, even in adults. At least until the evidence is gathered to resolve the issue, it seems premature to assume that children's phonological knowledge is any more abstract than their syntactic knowledge. The slot-and-frame schemas provide a useful alternative unit of analysis for children's fluency because their early phonological knowledge may well be tied to these if it is not abstract. It is worth mentioning that the authors do not reject the possibility that there are influences other than at the syntactic level that could influence fluency (e.g. prosodic planning), but the most detailed UB work to date with children concerns syntactic patterns. The argument is that until more is known about other forms of processing, the slot-and-frame schemas constitute a more empirically justified starting point for researching the planning units relevant to fluency in childhood than do the units of the generative approach.

How might schemas and operations interact online to affect the fluency of speech production? Naturalistic analysis should be expected to reveal a correlation between patterns of fluency and the location of schema boundaries as Howell et al. (1999) found with PW (c.f. Au-Yeung, Gomez & Howell, 2003, for Spanish; Dworzynski, Howell & Natke, 2003, for German). To date, the distribution of disfluencies in schemas has not been investigated, either with CWNS or CWS. It would be informative to investigate the validity of this hypothesised correlation for several reasons: 1) It would test directly the strength of the schema theory of early grammar as a psychologically real mechanism of language production; 2) A schema is an empirically identifiable mode of linguistic organisation for children and therefore constitutes an interesting candidate for investigating how fluency relates to syntax in a psychologically real sense. As mentioned above, the key point is that more disfluency would be expected at sensitive planning moments. The next section will make concrete suggestions for future research along these lines.

7. Suggestions for Future Fluency Research within the UB Approach

Through the course of this article, we have explored how the insights from the UB approach to language acquisition could relate to the existing research on fluency behaviour across development. There is much scope for reciprocal benefit by combining the efforts of researchers in child language and fluency. The UB approach has a tradition of embracing research from diverse disciplines (e.g. linguistics, psychology, neural net modeling) and it is hoped by the authors of the current paper that this

process of integration will continue in future by encompassing the area of fluency research and online speech production mechanisms. Several of the possible directions are suggested below.

The first logical step towards tying the two fields together would be to assess whether the units of speech (schemas) that Lieven et al. (2003) identified in Annie's speech constitute online planning units in speech production for Annie herself. Though Annie is a fluent child, CWS and CWNS show the same forms of disfluency (see section 4), so the units she used would be a useful basis for searching for similar planning units in the speech of CWS. This would entail investigating whether the boundaries of the schemas correlated with her patterns of fluency, namely whether moments of disfluency coincided with the point in time at which an operation would be required to fill a slot (i.e. immediately before the slot word). The initial signs are that Annie did show more disfluency at 3;0 than at 2;0 but there is a real need to study this systematically. If the operations on schemas were empirically verified as planning units in speech production, the analysis of Lieven and colleagues could be used to study CWS. Around 25-30 hours of data would be required from a CWS so that the utterances in the final session could be compared with the earlier utterances. The operations and schemas that were identified in this way would be expected to correlate with the fluency patterns of the CWS.

This type of research would consolidate and extend the UB findings by providing a different kind of evidence for the existence of schemas and by investigating their online role. It would also improve fluency research by providing an empirically justified unit of analysis rather than the existing ones that are based on traditional generative theory, which does not match the data on child language. To date, the PW adopted by Howell and colleagues has provided the best fit with the evidence on fluency compared with traditional grammatical units, but this is a unit motivated by phonological theory rather than being empirically derived. The 'slot-and-frame' operations of Lieven et al. (2003) are compatible with the existing data because most of the abstracted 'slots' in Lieven et al.'s schemas from Annie's data are content words that are preceded by function words at the end of the 'frame', for example Annie's '*Where's the X*?' could become '*Where's the bus*?' or '*Where's the cat*?' and her '*I want a Y'* could become '*I want a <u>toastie</u>'* or '*I want a <u>biscuit</u>'*. There is also work by Strenstrom and Svartvik (1994) on fluent adult English speakers that showed more fluency breakdown occurred on pronouns that were produced before verbs than after verbs (3.39% and 0.14% were repeated, respectively). The finding is consistent with the hypothesis that the PW is the unit of speech planning (i.e. predicts that function words will be repeated prior to content words rather than after them) but also that low-scope schemas are the units of speech planning, especially as these are often based around pronouns (see section 2). It would be interesting to compare traditional grammatical segmentation with PW segmentation and UB schema segmentation in terms of how well they predict the planning unit boundaries of children's speech as determined by their patterns of fluency and disfluency.

Another interesting issue that has been mentioned briefly is whether there are any differences in the syntactic knowledge or abilities of CWS compared with CWNS. The tools of the UB approach could be used to address this by assessing the syntactic knowledge of CWS (e.g. the novel word paradigm) and the nature of the naturalistic productions of CWS (e.g. how much abstraction versus lexical specificity is evident at 2 years compared with CWNS). The need to study this is supported by Silverman and Bernstein-Ratner's (2002) recent finding that there may be differences between CWS and CWNS in terms of the variety of lexical items they use in their speech, with CWS exhibiting less lexical variety than CWNS. It now remains to link this rather gross measure to the specific structures the children use, to see whether CWS may lag behind in the process of forming abstractions and so endure a more prolonged 'trade-off' period during which speech disfluencies occur. One problem with providing a rich data set from a young child who stutters is that data tend to be available at a later age, as it is only then that the child is identified as being a child who stutters. Given that UB theories of grammar propose that schemas or constructions underlie grammar right up to adulthood, however, the answer would be to adapt the analysis to search for more abstract level schemas than Lieven et al. (2003) did. Once such analyses were available, they would present a rich seam of data, but in the absence of a rich enough data set, an alternative less labour-intensive solution is possible. Carefully selected 'probe' items could be used. Instead of performing a full distributional analysis to determine the candidates for planning units, the existing evidence from adult data could be used to provide high frequency probe items such as Bybee and Scheibman's (1999) '*don't*' (mentioned earlier) to assess whether these correlated with patterns of fluency. Again, this type of research would both help to test the UB approach and shed further light on stuttering.

One more area in which the UB approach would benefit fluency research is in assessing the underlying phonological representations of CWS (though this has not been the focus of the current article). This would be a method of exploring the discrepancy between the fluency research which argues that the relevant planning unit for fluency behaviour is the PW, and the UB account, which suggest that low-scope schemas may be the units of planning in speech production for children. The issue could be approached empirically by examining lexical specificity in relation to phonology in a

similar way to that which UB researchers have done for syntax. If the lexical content of certain prosodic patterns in children's speech varies, then it could be claimed that the child has abstract prosodic knowledge, whereas if the same phrases or schemas always occurred with the same prosody, then the child could be argued to have lexically-specific prosodic knowledge instead. The UB approach would predict that early in life these prosodic and syntactic features are stored together in a schema, and only later do they become useable separately. A different issue concerning phonology would be to establish whether CWS have accurate and intact phonological representations compared to CWNS. A priming paradigm could be used, for example, to test whether CWS and CWNS who hear '*ba*' would be primed by this to name a picture of a banana more quickly. If so, then we could conclude that both CWS and CWNS possessed the full representation of *banana* and CWS do not stutter as a consequence of poor phonological representations. On top of the priming effect, if CWS were found to name the words slower than CWNS, then it could be argued that they encode phonological representations more slowly (as suggested by Kolk and Postma's 1997 'Covert Repair Hypothesis' account of stuttering), though this would not be the prediction of the EXPLAN model discussed earlier. Thus there are many ways in which the tools of the UB approach could be of benefit to fluency researchers interested in phonology.

To conclude, the UB approach to child language development and research on fluency have much to contribute to one another. The ideas presented here are by no means an exhaustive representation of what is possible. It would be very productive for both fields if researchers were to collaborate and combine the insights of both. At a practical level, deeper understanding of fluency development in general is essential for the proper management of the disorder in childhood and throughout life.

Acknowledgement. The first author was supported by the Wellcome Trust. The second author is grateful to Heike Behrens, Stephanie Brosda and Michael Tomasello for initial discussions on this topic.

References

Abbot-Smith, K., Lieven, E,. & Tomasello, M. (2001). 'What preschool children do and do not do with ungrammatical word orders'. *Cognitive Development*, **16**, 1-14.

Akhtar, N. (1999). 'Acquiring basic word order: Evidence for data-driven learning of syntactic structure'. *Journal of Child Language*, **26**, 339-56.

Akhtar, N., & Tomasello, M. (1997). 'Young children's productivity with word order and verb morpholgy'. *Developmental Psychology*, **33**, 952-965.

Allen, S. (1996). 'Aspects of argument structure in Inuktitut'. Amsterdam: John Benjamins.

Andrews, G., Craig, A., Feyer, A., Hoddinott, S., Howie, P., & Neilson, M. (1983). 'Stuttering: A review of research findings and theories circa 1982'. *Journal of Speech and Hearing Disorders*, **48**(3), 226-246.

Au-Yeung, J., Vallejo Gomez, I., & Howell, P. (2003). 'Exchange of disfluency from function words to content words with age in Spanish speakers who stutter'. *Journal of Speech, Language and Hearing Research*, **46**, 754-765.

Au-Yeung, J. Howell, P., & Pilgrim, L. (1998). 'Phonological words and stuttering on function words'. *Journal of Speech, Language and Hearing Research*, **41**, 1019-1030.

Bernstein, N. (1981). 'Are there constraints on childhood disfluency?'. *Journal of Fluency Disorders*, **6**, 341-350.

Bernstein-Ratner (1997). 'Stuttering: A psycholinguistic perspective'. In R. Curlee & G. Siegel, (eds.), *Nature and treatment of stuttering: New directions* (2nd edition). Needham, MA: Allyn & Bacon.

Blood, G., & Hood, S. (1978). 'Elementary school-aged stutterers' disfluencies during oral reading and spontaneous speech'. *Journal of Fluency Disorders*, **3**, 155-165.

Bloodstein, O. (1995). *A handbook on Stuttering* (5th edition). San Diego, CA: Singular Publishing Group, Inc.

Bloodstein, O., & Gantwerk, B. F. (1967). 'Grammatical function in relation to stuttering in young children'. *Journal of Speech and Hearing Research*, **10**, 786–789.

Bloodstein, O., & Grossman, M. (1981). 'Early stutterings: Some aspects of their form and distribution'. *Journal of Speech and Hearing Research*, **24**, 298–302.

Bock, K. (1986). 'Syntactic persistence in language production'. *Cognitive Psychology*, **18**, 355-387.

Borer, H., & Wexler, K. (1987). In E. Williams and T. Roeper (eds.), *Parameter Setting*. Reidel Publishing: Dordrecht, p. 123-172.

Bowerman, M. (1982). 'Starting to talk worse: Clues to language acquisition from children's late speech errors'. In S. Strauss (ed.), *U-shaped Behavioural Growth*. New York: Academic Press.

Braine, M. D. S. (1976). 'Children's first word combinations'. *Monographs of the Society for Research in Child Development*, **41** (1), Serial No. 164.

Braine, M. D. S. (1988). 'Modeling the acquisition of linguistic structure'. In Y. Levy, I. M . Schlesinger & M. D. Braine, (eds.), *Categories and processes in language acquisition*'. Hillsdale, NS: Lawrence Erlbaum Associates.

Brutten, G., & Hedge, M. (1984). 'Stuttering: A clinically-related overview'. In S. Dickson, (ed.), *Communication Disorders*. Glenview, IL: Scott Foresman.

Bybee, J. (1998). 'The Emergent Lexicon'. *Papers from the Regional Meetings, Chicago Linguistic Society*, **2**, 421-435.

Bybee, J., & Hopper, P. (2001) (eds.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam; Philadelphia: John Benjamins.

Bybee, J., & Scheibman, J. (1999). 'The effect of usage on degrees of constituency: The reduction of don't in English'. *Linguistics*, **37** (4), 575-596.
Bybee, J., & Thompson, S. (1998). 'Three frequency effects in syntax'. Paper presented at Twenty-third Annual Meeting of the Berkeley Linguistics Society.
Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). 'A construction based analysis of child directed speech'. *Cognitive Science*, **27**(6), 843-873.
Chang, F., Dell, G., Bock, K., & Griffin, Z. (2000). 'Structural priming as implicit learning: A comparison of models of sentence production'. *Journal of Psycholinguistic Research*, **29** (2), 217-229.
Childers, J., & Tomasello, M. (2001). 'The role of pronouns in young children's acquisition of the English transitive construction'. *Developmental Psycholgy*, **37**, 739-748.
Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
Culicover, P. (1999). *Syntactic Nuts*. Oxford: Oxford University Press.
Chang, F., Dell, G., Bock, K., & Griffin, Z. (2000). 'Structural priming as implicit learning: A comparison of models of sentence production'. *Journal of Psycholinguistic Research*, **29**(2), 217-229.
Conture, E. G. (1990). *Stuttering.* Englewood Cliffs, NJ: Prentice-Hall.
Davis, B., & MacNeilage, P. (1995). 'The articulatory basis of babbling'. *Journal of Speech and Hearing Research*, **38**(6), 1199-1211.
Dworzynski, K., Howell, P., & Natke, U. (2003). 'Predicting stuttering from linguistic factors for German speakers in two age groups'. *Journal of Fluency Disorders*, **28**, 95-113.
Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethik, S. (1994). 'Variability in early communicative development'. *Monographs of the Society for Research in child Development*, **59** (5), Serial #242.
Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology***, 31**, 41-81.
Fisher, C. (2000) *Who's blicking whom?: Word order in early verb learning.* Poster presented at the 11th International Conference on Infant Studies, Brighton, England.
Fisher, C. (2002) Structural limits on verb mapping: The role of abstract structure in 2.5-year-old's interpretations of novel verbs. *Developmental Science*, **5**, 55-64
Gaines, N. Runyan, C., & Meyers, S. (1991). 'A comparison of young stutterers' fluent versus stuttered utterances on measures of length and complexity'. *Journal of Speech and Hearing Research*, **34**, 37-42.
Gathercole, V., & Williams, K. (1994). 'Review of A. Radford, Syntactic theory and the acquisition of English syntax: the mature of early child grammars in English'. *Journal of Child Language*, **21**, 489-516.
Gee, J. P., & Grosjean, F. (1983). 'Performance structures: A psycholinguistic and linguistic apprasial'. *Cognitive Psychology*, **15**, 411–458.
Gershkoff-Stowe, L. (2002). 'Object naming, vocabulary growth, and the development of retrieval abilities'. *Journal of Memory and Language*, **46**, 665-687.
Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldberg, A. (1999). 'The emergence of the semantics of argument structure constructions'. In B. MacWhinney, (ed.), *The Emergence of Language*. Mahwah, NJ: Erlbaum, p. 197-212.
Gordon, P., & Luper, H. (1989). 'Speech disfluencies in nonstutterers: Syntactic complexity and production task effects'. *Journal of Fluency Disorders*, **14**, 429-445.
Hopper, P. (1987). 'Emergent Grammar'. *BLS* **13**, 139-157.
Howell, P. (2004). 'Assessment of some contemporary theories of stuttering that apply to spontaneous speech'. *Contemporary Issues in Communicative Sciences and Disorders*, **39**, 122-139.
Howell, P., & Au-Yeung, J. (1995). 'Syntactic determinants of stuttering in the spontaneous speech of normally fluent and stuttering children'. *Journal of Fluency Disorders* , **20**, 317-330.
Howell, P., & Au-Yeung, J. (2002). 'The EXPLAN theory of fluency control and the diagnosis of stuttering'. In E. Fava, (ed.), *Current Issues in Linguistic Theory series: Pathology and therapy of speech disorders*. Amsterdam: John Benjamins, pp. 75–94
Howell, P., Au-Yeung, J., & Sackin, S. (1999). 'Exchange of stuttering from function words to content words with age'. *Journal of Speech, Language and Hearing Research*, **42**, 345-354.
Howell, P., Au-Yeung, J., & Sackin. S. (2000). 'Internal structure of content words leading to lifespan differences in phonological difficulty in stuttering'. *Journal of Fluency Disorders*, **25,** 1-20.
Howell, P., Davis, S., & Au-Yeung, J. (2003). 'Syntactic development in fluent children, children who stutter, and children who have English as an additional language'. *Child Language Teaching and Therapy*, **19**, 311-337.
Jackendoff, R. (1996). *The architecture of the language faculty*. Bradford: MIT Press.
Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
Jakielski, K. J. (1998). *Motor organization in the acquisition of consonant clusters.* Dissertation / PhD thesis: University of Texas, Austin.
Kadi-Hanifi, K., & Howell, P. (1992) 'Syntactic analysis of the spontaneous speech of normally fluent and stuttering children'. *Journal of Fluency Disorders*, **17**, 151-170.
Karniol, R. (1995). 'Stuttering, language, and cognition: A review and a model of stuttering as suprasegmental sentence plan alignment (SPA)'. *Psychological Bulletin*, **117**(1), 104-124.
Kolk, H., & Postma, A. (1997). 'Stuttering as a covert repairs phenomenon'. In R. F. Curlee & G. M. Siegel (eds.), *Nature and treatments of stuttering: New directions* (pp. 182–203). Needham Heights, MA: Allyn & Bacon.

Langacker, R. (1987). *Foundations of Cognitive Grammar, Vol.1*: Theoretical prerequisites. Stanford: Stanford University Press.

Langacker, R. (1991). *Foundations of Cognitive Grammar, Vol. 2*: Descriptive application. Stanford: Stanford University Press.

Lebeaux, D. (1988). 'Language acquisition and the form of grammar'. Unpublished dissertation, University of Massachussetts.

Levy, Y. (1983). 'The acquisition of Hebrew plurals: The Case of the Missing Gender Category'. *Journal of Child Language*, **10**(1), 107-121.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). 'Early syntactic creativity: A Usage-Based approach'. *Journal of Child Language*, **30**, 1-38.

Lieven, E., Pine, J., & Baldwin, G. (1997). 'Lexically-based learning and grammatical development'. *Journal of Child Language*, **24**, 187-220.

Logan, K., & Conture, E. (1997). 'Selected temporal, grammatical, and phonological characteristics of conversational utterances produced by children who stutter'. *Journal of Speech, Language and Hearing Research*, **40**, 107-120.

MacNeilage, P., & Davis, B. (1990). Acquisition of speech production: Frames, then content. In M.Jeannerod (ed.), *Attention and Performance XIII: Motor representation and control.* Hillsdale NJ: Erlbaum.

Marchman, V., & Bates, E. (1994). 'Continuity in lexical and morphological development: A test of the critical mass hypothesis'. *Journal of Child Language*, **21**, 339-366.

Maratsos, M., Gudeman, R., Gerard-Ngo, P., & DeHart, G. (1987). 'A study in novel word learning: the productivity of the causative'. In B. MacWhinney (ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

Naigles, L. (1990). 'Children use syntax to learn verb meanings'. *Journal of Child Language*, **17**, 357-374.

Ono, T., & Thompson, S. (1995). 'What can conversation tell us about syntax?' In P. Davis (ed.), *Alternative linguistics: Descriptive and theoretical models*. Amsterdam, Netherlands: John Benjamins Publishing Co, p. 213-271.

Pickering, M., and Branigan, H. (1998). 'The representation of verbs: Evidence from syntactic priming in language production'. *Journal of Memory and Language*, **39**, 633-653.

Pickering, M., & Branigan, H. (1999). 'Syntactic priming in language production'. *Trends in Cognitive Science*, **3** (4), 136-141.

Pickering, M., Branigan, H., Cleland, A., & Stewart, A. (2000). 'Activation of syntactic information during language production'. *Journal of Psycholinguistic Research*, **29** (2), 2000.

Pine, J., & Lieven, E. (1993). 'Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech'. *Journal of Child Language* **20**, 551-571.

Pine, J., Lieven, E., & Rowland, C. (1998). 'Comparing different models of the English verb category'. *Linguistics*, **36** (4), 807-830.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, Mass.: Harvard University Press.

Pinker, S. (1989). *Learnability and Cognition*. Cambridge, Mass.: MIT Press.

Pinker, S. (1995). 'Language acquisition'. In L. R. Gleitman, M. Liberman, & D. N. Osherson (eds.), *An Invitation to Cognitive Science*, *2nd edition*. Vol. 1: Language. Cambridge, MA: MIT Press.

Pinker, S., Lebeaux, D., & Frost, L. (1987). 'Productivity and constraints in the acquisition of the passive'. *Cognition*, **26**, 195-267.

Pizutto, E., & Caselli, C. (1994). 'The acquisition of Italian verb morphology in a cross-linguistic perspective.' In Y. Levy (ed.), *Other children, other languages*. Hillsdale, NJ: Erlbaum.

Plunkett, K., & Marchman, V. (1993). 'From rote learning to system building: Acquiring verb morphology in children and connectionist nets'. *Cognition*, **48,** 21-69.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax: the nature of early child grammars in English*. Oxford: Blackwell.

Radford, A. (1995). 'Phrase structure and functional categories.' In P. Fletcher & B. MacWhinney (eds.), *Handbook of child language*. Oxford: Blackwell.

Radford, A. (1996). 'Towards a structure building model of acquisition'. In H. Clahsen (ed.), *Generative perspectives on language acquisition*. Amsterdam: John Benjamins.

Ratner, N. B. (1997). 'Stuttering: A psycholinguistic perspective'. In R. Curlee & G. Siegel (eds.), *Nature and treatment of stuttering: New directions* (2nd edition, pp. 99-127). Boston: Allyn & Bacon.

Ratner, N., & Sih, C. (1987). 'The effects of gradual increases in sentence length and complexity on children's dysfluency'. *Journal of Speech and Hearing Disorders*, **52**(3), 278-287.

Roberts, K. (1983). 'Comprehension and production of word order in Stage I'. *Child Development*, **54**, 443-449.

Roeper, T., & Williams, E. (eds.) (1987). *Parameter Setting*. Dordrecht: Reidel Publishing.

Rubino, R., & Pine, J. (1998). 'Subject-verb agreement in Brazilian Portugese: What low error rates hide'. *Journal of Child Language*, **25**, 35-60.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). 'Testing the abstractness of children's linguistic representations: lexical and structural priming of syntactic constructions in young children'. *Developmental Science,* **6** (5), 557–567.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (submitted to *Language Learning and Development*). 'Structural priming as implicit learning in language acquisition: The persistence of lexical and structural priming in 4-year-olds'.

Schenkein, J. (1980). 'A taxonomy for repeating action sequences in natural conversation'. In B. Butterworth (ed.), *Language Production (Vol. 1)* (p. 21-47). San Diego, CA: Academic Press.

Schilperoord, J. (1996). *It's about time. Temporal aspects of cognitive processes in text production*. Amsterdam: Rodopi.

Schilperoord, J. (1997). 'Temporele modificatie in clauses; een pause-analytische studie naar teksproductie'. In H. van den Bergh, D. M. L. Janssen, N. Bertens & M. Damen (eds.), *Taalgebruik ontrafeld. Bijdragen van het zevende VIOT-taalbeheersinscongres gehouden op 18, 19 en 20 december 1996 aan de Universiteit van Utrecht* (pp. 263-274). Dordrecht: ICG Publications.

Schilperoord, J., & Verhagen, A. (1998). 'Conceptual dependency and the clausal structure of discourse'. In J-P Koening (ed.), *Discourse and cognition. Bridging the gap* (pp. 141-163). Stanford: CSLI Publications.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Silverman, S., & Bernstein-Ratner, N. (1997). 'Syntactic complexity, fluency, and accuracy of sentence imitation in adolescents'. *Journal of Speech, Language, and Hearing Research*, **40**, 95-106.

Silverman, S., & Bernstein-Ratner, N. (2002). 'Measuring lexical diversity in children who stutter: application of *vocd*'. *Journal of Fluency Disorders*, **27**, 289-304.

Strenstrom, A.B., & Svartvik, J. (1994). 'Imparsable speech: Repeats and other nonfluencies in spoken English'. In N. Oostdijk & P. de Haan (eds.), *Corpus-based research into language.* Amsterdam: Rodopi.

Theakston, A., Lieven, E., Pine, J., & Rowland, C. (2001). 'The role of performance limitations in the acquisition of verb-argument structure: An alternative account'. *Journal of Child Language*, **28**(1), 127-152.

Tomasello, M. (1992). *First Verbs: a case study of early grammatical development*. Cambridge, England: Cambridge University Press.

Tomasello, M. (2000). 'Do young children have adult syntactic competence?' *Cognition*, **74** (3), 209-253.

Tomasello, M. (ed.) (2003). *The new psychology of language: cognitive and functional approaches to language structure, Volume* 2. Mahwah, NJ: Lawrence Erlbaum Assc.

Tomasello, M., & Brooks, P. (1998). 'Young children's early transitives and intransitive constructions'. *Cognitive Linguistics*, **9**, 379-395.

Valian, V. (1986). 'Syntactic categories in the speech of young children'. *Developmental Psychology,* **22**, 562-579.

Valian, V. (1991). 'Syntactic subjects in the early speech of American and Italian children'. *Cognition,* **40**, 21-81.

Verhagen, A. (2001). 'Subordination and discourse segmentation revisited, or: Why matrix clauses may be more dependent than complements'. In E. Sanders, J. Schilperoord and W. Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 337-357). Amsterdam: John Benjamins.

Vogel Sosa, A. & MacFarlane, J. (2002). 'Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of'. *Brain and Language*, **83**, 227-236.

Wall, M. (1980). 'A comparison of syntax in young stutterers and nonstutterers'. *Journal of Fluency Disorders*, **5**, 345-352.

Wall, M., Starkweather, C., & Cairns, H. (1981). 'Syntactic influences on stuttering in young child stutterers'. *Journal of Fluency Disorders*, **5**, 345-352.

Watkins, R. V., & Yairi, E. (1997). 'Language production abilities of children whose stuttering persisted or recovered'. *Journal of Speech, Language, and Hearing Research*, **40**, 385-399.

Westby, C. (1974). 'Language performance of stuttering and nonstuttering children'. *Journal of Communications Disorders*, **12**, 133-145.

Wingate, M. E. (1988). *The structure of stuttering*. New York: Springer-Verlag.

Wingate, M. (2002). *Foundations of stuttering.* New York: Academic Press.

Yaruss, S. (1999). 'Utterance length, syntactic complexity, and childhood stuttering'. *Journal of Speech, Language, and Hearing Research*, **42**, 329-344.

RESEARCH COMMENTARIES ON SAVAGE AND LIEVEN

Commentary on 'Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?' by C. Savage and E. Lieven

Lisa Gershkoff-Stowe
Department of Speech and Hearing Sciences
Indiana University
gershkof@indiana.edu

Abstract. Savage and Lieven's (2004) usage-based approach to developmental stuttering is examined in light of key concepts of dynamic systems theory.
Keywords: Developmental stuttering, dynamic systems theory, usage-based theory.

1. Introduction

The purpose of this commentary is to consider childhood disfluency from a dynamic systems perspective. By this view, disfluency is conceptualized as the transient property of a system undergoing change. I begin by describing briefly some basic assumptions of dynamic systems theory and consider how the model lends itself to investigating problems of fluency control. Next, I contrast aspects of the model to Savage and Lieven's (2004) usage-based approach. I then conclude by offering intuitions about key parameters for future study.

2. A Dynamic Systems View

Several decades of research on language in normal adult speakers have revealed the enormous complexity of the processing system that underlies speech. Numerous integrated systems comprise our ability to formulate, retrieve, and produce syntactical constructions of even the simplest kind. Among the relevant system variables that are linked to fluency are syntax, phonology, and the lexicon. In addition, several factors related to the circumstance of speaking are known to affect fluency. These variables operate in the flow of continuous time and include, for example, rate of speaking, temporal planning, and variations in context.

Dynamic models are particularly appropriate for studying complex patterns of fluency and disfluency because of their emphasis on behavioral variability as an index of change. At certain points of instability, it is often possible to specify the parameters that disrupt the system and drive it to new states of stability. Such parameters can then be manipulated using experimental methods to test causal predictions. In dynamic systems theory, attention is given to the collective activity of the many subcomponents that comprise the entire system. Together these components form a stable configuration of behavior under certain conditions and generate new patterns of behavior under others. Each component, moreover, has a unique developmental trajectory and differential rate of growth. This means that any one component can serve as a critical or rate-limiting factor that constrains the range of tasks to which the individual is capable of responding. A major concern with respect to language, then, is how fluent speech is affected by the simultaneous and mutually constraining interactions of system variables that are themselves changing over time.

Dynamic systems theory considers such long-term changes as they occur in ontogenetic time, but in addition, views the moment-to-moment processes that take place in real time. Both levels are essential to explanations of development; both operate to shape behavior in mutually influencing and continuously changing ways. Thus at any given moment, patterns of fluency reflect the performance biases, or preferred states, of a system as they emerge within the proximate here-and-now effects of a particular task-context. Importantly, these biases are age- and experience-dependent. The value of dynamic systems as an explanatory model for development is its potential to incorporate many diverse factors into one coherent theory of change.

3. Contrasting Models

There appears to be much common theoretical ground between usage-based theory and dynamic systems theory. As such, Savage and Lieven (2004) have provided a strong foundation for approaching the problem of developmental fluency from a dynamic systems perspective. In particular, the authors focus directly on questions that concern key points of transition: Why do children show a temporary increase in disfluency between two and three years of age? Why do some children continue to stutter while others progress toward adult-like patterns of fluent speech? Because children are especially

sensitive to perturbations at times of transition, it may be possible to identify the constraints and catalysts that move the system from regions of stability to instability. For example, Savage and Lieven (2004) hypothesize that "stuttering prevails at points of vulnerability in language development when the system is under strain through the acquisition of a linguistic skill" (**pg. 1**). As children adapt to the changing demands placed on the language processing system, they should show marked improvements in fluency.

Also consistent with a dynamic systems view, Savage and Lieven have abandoned traditional distinctions between competence and performance and consider, instead, the circumstances of online processing. Thus their usage-based model recognizes that grammar is dynamic and experience-driven. However, important differences exist with respect to the nature of linguistic representation. In a dynamic model, knowledge is conceptualized as the emergent product of local processes - those that occur in the moment-to-moment activity of real-time. Knowledge by this view is fluid and probabilistic; it depends on the coupling and uncoupling of multiple systems involving the present context and recent past history of the system. In contrast, Savage and Lieven characterize knowledge as lexically-specific schemas that elaborate into more abstract categories. Accordingly, a primary question in usage-based theory is whether or not children are in full possession of a particular structure (pg. 6). A more useful depiction, however, is one in which knowledge is at all times softly-assembled in a time-dependent and context-specific manner. This means that children's underlying representations exist not as abstractions but only as responses to task demands that can be potentially known by the researcher.

4. Conclusion

I have suggested that a dynamic systems approach can offer insight into the processes of developmental stuttering beyond traditional linguistic explanations. This is accomplished through consideration of the language processing system as a product of the dynamic interplay of many contributing subsystems. Within this view, patterns of fluency emerge from the joint effects of both real and developmental time.

Dynamic systems theory suggests a general strategy for addressing the underlying mechanisms of stuttering that involves a component-by-component analysis of the individual subsystems that comprise the total language processing system. Savage and Lieven have taken important steps toward achieving this goal by studying children's online production. Dynamic systems theory, like usage-based theory, maintains that syntax, semantics, and phonology are intricately connected. Importantly, however, not only linguistic events, but also domains such as working memory, attention, retrieval, and motor processes may contribute to changes in patterns of fluency. Strand (1992), for example, has suggested that disfluency rates may be tied to the automaticity of the language system. He notes that to process language in a highly reflexive way, children require extended effort and practice in speaking. Finally, at the local level, dynamic systems theory suggests that fluency breakdown may result from a number of different factors that function to increase processing demands on the immature speaker. Much attention has been rightly given to syntactic complexity; other variables, such as utterance length and prosodic planning, may also play a role.

By framing the problem of childhood disfluency in dynamic systems terms, researchers may be in a position to capture the complexity of development as it changes over multiple levels of analysis and different scales of time. This knowledge, in turn, should provide new insights into the mechanisms underlying normal speech production.

References

Savage, C., & Lieven, E. (2004). Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering? *Stammering Research*, **1**, 83-100.

Strand, E. A. (1992). The integration of speech motor control and language formulation in models of acquisition. In R. S. Chapman (Ed.), *Processes in language acquisition and disorders* (pp. 86-107). St Louis, MO: Mosby-Year Book.

Thelen, E. & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Commentary on 'Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?' by C. Savage and E. Lieven

O. P. Skljarov
Research Institute of ETN & Speech, 190013, St.Petersburg, Bronnitskaja, 9, Russia
skljarov@admiral.ru

Abstract. Savage and Lieven's (2004) usage-based approach to developmental stuttering is examined in light of key concepts of dynamic systems theory.
Keywords: Developmental stuttering, dynamic systems theory, usage-based theory.

1. Introduction

One of the main points in Savage and Lieven's (2004) paper "Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?" is a comparison of the usage-based psycho-linguistic approach of the authors (UB model) with the phonological word approach offered by Howell (20041, b) and Howell and Au-Yeung (2002) (the PW model). The relevance of the UB model to fluency research is that it should be able to generate concrete predictions about where disfluency is most likely to occur (i.e. in which forms of construction and in which parts of those constructions). The UB model uses a psycho-linguistic approach that claims that the basic cause of development stuttering is associated with the abstract slots in grammatical strings (Savage & Lieven, 2004).

Consistent with the UB model, the more frequently used and more concrete phrases only require a direct and automated retrieval step, whereas more creative productions require utterance construction in which certain abstract slots are filled with lexical items. The moments in speech that precede the more abstract parts of an utterance will be more vulnerable to disfluency as a consequence of the greater demands placed on the language production system at such points, as preparation occurs for the subsequent part of the utterance. The loci of these vulnerable moments will vary for adults and children, because children's representations change gradually over time (Savage & Lieven, 2004).

However, the steps in the above argument are, as yet, unsubstantiated. The empirical evidence to date for UB models has focused mainly on the nature of the underlying linguistic representations that children and adults possess and not how they use these to produce speech. It is still an open question as to whether the UB models' representations have any real psychological significance in online speech production. The authors do tie the theory with the ideas of Gee and Grosjean (1983) as Howell and Au-Yeung (2002) did previously in their PW model. Gee and Grosjean (1983) did not directly examine functionally derived units; their units of analysis were 'prosodic bundles' that were derived from an algorithm based on formal linguistic principles. An important characteristic was that these bundles did not correlate with syntactic boundaries. In fact they closely approximated the 'phonological word' unit used by Howell and Au-Yeung (2002). In sum, tying UB theory in with Gee and Grosjean's work allows Savage and Lieven to predict that words that are stuttered would be expected to occur at the start of processing boundaries and these points may not be the traditional grammatical boundaries that occur in formal linguistics.

There are several reasons that have been given in the literature as to why stuttering in childhood is 'anomalous' (i.e. why it occurs on simple, rather than complex, word types). Wingate (2002) argued that the pattern of stuttering observed in childhood is just normal nonfluency, not stuttering. This would also explain why there is a high rate of recovery from childhood 'stuttering'. The majority of authorities maintain that there is developmental change in the pattern of stuttering (e.g. Conture, 1990). Howell (2004) goes further and argues, from his phonological approach, that the different patterns of stuttering between childhood and adulthood are two distinct ways of dealing with situations where there is material that is difficult to prepare in the speech context. The childhood form of stuttering links in with Yairi's view that maintains word repetitions are stuttering-like disfluencies that precede difficult words. Howell's approach is more elaborate and argues that the childhood form of disfluency is associated with getting the words ready in time.

The problems with the phonological approach are shown in the study by Throneburg, Yairi, Paden (1994). These authors examined the effect of consonants of various types classified into whether they occurred early or late in development, by the number of syllables in the word and presence or not of consonant strings. The children they used were divided also into different classes depending on phonological ability and severity of stuttering. The authors concluded, that phonological difficulty of the disfluent word, and the fluent word following it, did not contribute to fluency breakdown regardless

of the childrens' phonological ability and stuttering severity. On the basis of these observations the authors suggest that perhaps phonological difficulty is not a sensitive indicator of the motor demand placed on the speech apparatus. This conclusion also has implications for Postma, Kolk and Povel's (1990) assertion that speech difficulties of children who stutter may result from problems with the central premotor planning of the speech act. If young children who stutter (with or without accompanying phonological problems) have a general speech-planning problem, as suggested by Postma et al. (1990), Throneburg et al.'s (1994) results imply that this problem is not aggravated when words are phonologically more difficult. Thus, the supposed planning problems must be reflected at some other level of speech processing.

Throneburg et al. (1994) also maintain that, to the extent that phonological difficulty reflects the complexity of articulatory gestures, the findings do not seem to support the motor discoordination hypothesis of stuttering (MacKay & MacDonald, 1984; St. Louis, 1991; van Riper, 1982). The essence of the latter authors views is that fluency disorders arise because of the way phonologically difficult expressions (i.e., the expressions with complex sound structure and, maybe, with complex rhythm), are created at the cognitive level that makes their lower level motor programs difficult to generate.

There is an interesting alternate hypothesis offered by Wingate (1988). According to this view, fluency disorders can arise because of the increased rhythmic demands placed on expressions with an irregular rhythm when they access their lexeme forms. However comparative examinations that have been carried out between adult speakers who stutter and fluent speakers have not confirmed (nor refuted) this hypothesis (Hubbard & Prins, 1994). Given this ambivalence, we undertook tests of Wingate's hypothesis in studies of speech rhythm in speakers who stutter.

Skljarov (2004) demonstrated a trend to simplification of rhythms in speakers who stutter, i.e. a change from irregular to regular temporal structure is observed in these speakers. This outcome is consistent with Wingate's (1988) hypothesis, van Riper's (1982) observations about stuttering in adults, with Howell's (2004) experimental data on children who stutter and Savage and Lieven's (2004) views about children who stutter. Stuttering in children frequently shows in tonoclonic form. Features associated with perseveration only appear in the chronic adult form. The co-occurrence of these two forms of stuttering (tonoclonic forms and perseveration) in adults may explain why Hubbard and Prins (1994) failed to confirm Wingate's hypothesis.

Howell's (2004) prediction that we can expect words a) to be stuttered at processing boundaries and b) these may not be the traditional grammatical boundaries of formal linguistics is also consistent with our point of view. Our view a) uses Feigenbaum's that fluency breakdown occurs at the transitions between synharmonic and syllabic branches, and a propos of b), Skljarov (2004) notes that these may not be the traditional grammatical boundaries of formal linguistics.

It is of particular note that our point of view is consistent with Savage and Lieven's position concerning segment boundaries and how they vary over developmental time as children build up more abstract representations (see Skljarov, 2004). However, we disagree Savage and Lieven's statement that the problem is not resolved because, given the UB approach is still young. They consider that the UB approach has focussed mainly on the nature of the underlying grammatical representations and consider that little is yet known about how the different routes to production (direct or via abstract units) and how they interact online. However, elsewhere we have offered the rhythm disorders theory which pinpoints that the rhythm route, along with the physical and physiological representations in the speech production system, leads to chaos which appears as stuttering (Skljarov, 1998a, 1998b, 1999, 2003a, 2003b, 2004). Thus, Savage and Lieven's statement that there is little evidence that directly assesses what the alternative units of processing may be, is incorrect as we have offered a scenario by which the rhythm route in speech leads to chaos that is manifest as stuttering.

We also do not agree with these authors' statement that no real theory has yet been provided from the UB perspective to explain the operation of online mechanisms involved in children's speech production. Such theory is available in the area of stuttering and has been applied to early language production and fluency (Skljarov, 2004).

References

Conture, E. G. (1990). *Stuttering.* Englewood Cliffs, NJ: Prentice-Hall.

Gee, J. P., & Grosjean, F. (1983). 'Performance structures: A psycholinguistic and linguistic apprasial'. *Cognitive Psychology*, **15**, 411–458.

Howell, P. (2004a). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Research*, **1**, 31-46.

Howell, P. (2004b). 'Assessment of some contemporary theories of stuttering that apply to spontaneous speech'. *Contemporary Issues in Communicative Sciences and Disorders*, **39**, 122-139.

Howell, P., & Au-Yeung, J. (2002). The EXPLAN theory of fluency control and the diagnosis of stuttering. In E Fava (Ed.), *Pathology and therapy of speech disorders* (pp. 75-94). Amsterdam: John Benjamins.

Hubbard, C.P., & Prins, D. (1994). Word familiarity, syllabic stress pattern, and stuttering. *Journal of Speech and Hearing Research*, **37**, 564-571.

MacKay, D., MacDonald, M. (1984). Stuttering as a sequencing and timing disorder. In: R. Curlee, W. Perkins (Eds.), *Nature and treatment of stuttering: New directions*. San Diego, CA: College-Hill.

Postma, A., Kolk, H.,& Povel, D. (1990). Speech planning and execution in stutterers. *Journal of Fluency Disorders*. **15**, 49-59.

Savage, C., & Lieven, E. (2004). Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering? *Stammering Research* **1**, 83-100.

Skljarov O.P. (1998a). Neurophysiological aspects of recurrent functioning of "hidden" variables in speech apparatus. *Zhurn. Vysh. Nervn. Deit*, **48**. 827-835.

Skljarov O.P. (1998b). Self-organizing nature of speech rhythm. *Biofizika*, **43**, 152-158.

Skljarov O.P. (1999). Nonlinear neurodynamics in repesentation of a rhythm of speech. *Journal of Biological Physics*. **25**. 223-234.

Skljarov O.P. (2003a). Biophysical basis of the universality principle at the speech perception. *Biofizika*, **48**, 553-557.

Skljarov O.P. (2003b). Lagrange formulation of self-organizing problem for neuron ensemble at the counting of non-linear dissipation of energy. *Biofizika*, **48**, 701-705.

Skljarov, O.P. (2004). Speech development and scenario of its V-rhythms' development. *Electron Journal "Technical Acoustics" http:/webcenter.ru/~eeaa/ejta* 2004, **7**.

St. Louis, K. The stuttering/articulation connection. In: H.Peters, W. Hulstijn, & C.Starkweather (Eds.). *Speech motor control and stuttering*. Amsterdam:Elsevier Science Publishers, 1991. pp. 393-400

Throneburg, R.N., Yairi, E., Paden, E. P. Relation between phonologic difficulty and the occurence of disfluencies in the early stage of stuttering. Journal of Speech and Hearing Research. 1994. Vol. 37. pp. 504-509.

Van Riper, C. (1982). *The nature of stuttering*. Englewood Cliffs. NJ: Prentice Hall.

Wingate M. (1988). *The structure of stuttering*. N. Y. Springer Verlag.

Wingate, M. (2002). *Foundations of stuttering.* New York: Academic Press.

Commentary on 'Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?' by C. Savage and E. Lieven

Dr Martin Schwarz
Director of The National Center For Stuttering
mfs3@nyu.edu

1. Commentary on Savage and Lieven

The article by Savage and Lieven (2004) "The Usage-Based Approach to Naturalistic Analysis of Developmental Stuttering" represents a coherent survey of language development in both normal speaking individuals and those who stutter. I have merely two questions. Since it is well known that the most typical form of the onset of stuttering is marked by "effortless" repetitions of sound, syllables or words at the beginning of sentences, would the authors care to comment precisely how their linguistic model would account for this exact behavior? Also, it has been documented that when children and adults who stutter speak to themselves out loud alone they are, in the main, totally fluent. Could the authors, using the model they have put forth, precisely explain this behavior?

Reference

Savage, C., & Lieven, E. (2004). Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering? *Stammering Research*, **1**, 83-100.

AUTHORS' RESPONSE TO COMMENTARIES
Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?

C. Savage

Department of Psychology, University College London, Gower St., London WC1E 6BT, England
c.savage@ucl.ac.uk

Abstract. The author offers responses to commentaries by Gershkoff-Stowe, Skljarov, and Schwarz. The overall response was positive to the argument for introducing the Usage-Based Approach to the study of childhood stuttering. The main points made in this response concern the role of non-syntactic processes and systems in leading to disfluency and the relationship between short and long term temporal influences on speech production. The author concludes by suggesting that it would be useful to continue the dialogue between the various sub-disciplines that are relevant to developmental stuttering. In particular, it would be useful to continue efforts to unite research on fluency with the most recent developments in child language.
Key Words: Usage-Based Approach, Dynamic Systems.

1. Introduction

I welcome the opportunity to respond to the commentaries offered by Lisa Gershkoff-Stowe, Oleg Skljarov and Martin Schwarz. It is pleasing that the response to the Savage and Lieven (2004) article is generally positive. Support is evident for the argument that it would benefit research into stuttering in childhood to adopt the Usage-Based (UB) approach that is used in the literature on fluent language development. The commentaries also raised some interesting questions and issues that either go beyond the original article or require clarification, particularly about the relationship of non-syntactic processes to fluency. Response is made to these and it is recommended that continuing dialogue takes place between these sub-disciplines.

2. Response to commentaries

Gershkoff-Stowe

Gershkoff-Stowe (2004) suggests that the UB approach shares much common theoretical ground with dynamic systems theory, and I am inclined to agree with this. Gershkoff-Stowe's comments are very interesting and suggest that much could be achieved by uniting the efforts of these approaches to tackle the questions of fluency early in life. The dynamic systems model emphasizes the relation between multiple systems, with an emphasis on the temporal dimension of speech production, in terms of both the short (recent or immediate precursors to speech) and long time-scales (speech experience across development). I agree that this is a crucial aspect of any theory of speech processing and makes the picture particularly complex and intricate for children, who lack much of the long term linguistic experience of adults and, conversely, are highly susceptible to short term influences because these constitute a higher proportion of their overall language input. It is of note, in fact, that much of the existing literature on childhood stuttering is guided by the models of language production that have been developed for the adult literature (e.g. Levelt, 1989) that have paid little attention to the possibility that children could process speech differently. This issue becomes particularly pertinent when one acknowledges that the generativist argument for a system of innate, universal linguistic knowledge does not stand up against the empirical evidence obtained from young children (as referred to in Savage & Lieven, 2004). If the knowledge components of the language production system are immature, then there is no reason to assume that the system functions in the same way as adults in terms of processing either.

One interesting challenge that can be expected to arise by giving the temporal aspects of language use a central position in a theory, is at what point the two timescales intersect. That is, at what point would a short term effect be better described as a long term feature of the system, or 'knowledge'. The issue can be highlighted by considering the difficulty encountered when one tries to define the nature of the syntactic priming effect for young children. In the adult literature, it was long considered that priming reflected a short term activation of static linguistic knowledge that subsides after only a few seconds (e.g. Bock & Loebell, 1990; Pickering & Branigan, 1999) but recently evidence has been gathered that suggests the effect is longer lived and may constitute a form of implicit learning (e.g. Bock & Griffin, 2000; Chang, Dell, Bock & Griffin, 2000). The complexity arises when we attempt to define more concretely where the cut off point exists between priming and learning, for example, for adults the effect can last across ten intervening sentences (Bock & Griffin, 2000), but it has not yet been established whether it could last across three days. If it did not, then would this constitute a

transient activation that lasted longer than initially thought, or would it be better defined as a short term learning phenomenon? Perhaps with repeated priming on day two after the effect, priming could endure until the third day by receiving a 'boost'. Would this still constitute a priming paradigm, spread over days, or would it constitute a learning paradigm? The question is a qualitative one. Is it justified or meaningful to argue that priming and learning are qualitatively different phenomena, or does the difference between the two reside in the fact that they occupy different ends of a continuum between short- and long-term learning as determined by the number and frequency of stimuli, that is a quantitative difference? The issue is more complex for children, because their linguistic experience prior to priming is far less than that of adults and as such is presumably more flexible and less stable. The priming stimuli would be expected to reach the required level to constitute learning at an earlier stage, because they are a more significant input variable than for adults in relation to the existing knowledge base. In fact, Savage, Lieven, Theakston and Tomasello (under review) recently conducted a study that addressed some of these questions and uncovered some of the features of syntactic priming in young children. They found that the priming effect for passives in 4-year-old fluent speakers lasted up to a month after the first presentation, but only if it was 'reinforced' after a week, by children being given opportunity to produce target items again at this intermediary stage. These findings support the hypothesis that priming constitutes a form of learning in children, but also raise new issues. One major issue is the extent to which the priming effect is maintained by becoming reciprocal, that is by the child 'self-priming'. A short term effect could become a longer term effect with no further deliberate external prompting if it set in motion a circular process. A child's response to priming could be to increase their own usage of a particular construction, which could then prompt others engaged in communication with them to use it more, and the increased usage of both child and interlocutor would influence the child's own productions to use the construction still more. The nature of the relationship between short- and long-lived effects on the language system, particularly in children, is very interesting and warrants further exploration. In short, I would agree with Gershkoff-Stowe's focus on the dynamic, temporal aspects of the language system.

Both Savage and Lieven (2004) and Gershkoff-Stowe (2004) acknowledge the need for researchers into childhood disfluency to integrate what we know about the nature of children's early knowledge with an exploration of how children perform the online process of speech production. The UB approach shares with the dynamic systems framework an emphasis on the 'dynamic' aspect of language development over the long term, with the former offering a thorough analysis of the changing nature of children's linguistic knowledge over time. The short term dynamics of the language production system also need to be considered, as these constitute the mechanistic half of the attempt to explain childhood stuttering. To this end, Savage and Lieven (2004) drew on the EXPLAN model (Howell & Au Yeung, 2002) because of its focus on the interface between language knowledge and processing. The dynamic systems model also could also provide insight along these lines by introducing a broader perspective on the relationship between the various different sub-systems involved in language production, including non-linguistic influences such as attention. This seems a relevant approach to take, in light of recent evidence from Bosshardt (2002) that German speaking adults who stutter have problems with executive control of their language production subsystems. In combination, the approaches mentioned offer a route by which to unite the research on linguistic knowledge and processing, and develop a more complete explanation of developmental stuttering. Such an explanation demands a more concrete model than is currently available for children of which types of processing come into play at which points in time during speech production and what the nature is of these across development. I would suggest that the best way to achieve this is to promote dialogue between researchers in all relevant sub-disciplines, from linguistics to computer science to psychology.

One comment that I would like to take up on is Gershkoff-Stowe's (2004) view on abstract knowledge. She claims that "a more useful depiction…[of abstract knowledge than that used be Savage & Lieven, 2004]… is one in which knowledge is at all times softly-assembled in a time-dependent and context-specific manner. This means that children's underlying representations exist not as abstractions but only as responses to task demands that can be potentially known by the researcher". Whilst I am sympathetic to her rejection of abstract knowledge as permanent and inflexible, this definition appears to lack a definition of what it means to possess any permanent capacities. We can argue that children lack something that allows adults to operate on language at an abstract linguistic level when they need to, even if their default would be to use more frequent concrete 'chunks', but if so we must define what it is that children lack. The UB approach continues to refer to the capacity behind this adult skill as 'abstract knowledge' and construes it as residing in connections within a network of stored linguistic experience. The 'abstract' knowledge, in this sense, exists only as connections between remembered exemplars (e.g. Bybee, 1998). The argument is that abstract knowledge refers to the capacity of adults to produce an utterance on the basis of more distant connections between stored elements of linguistic

experience should more unusual and creative sentences be required, in effect creating language based on abstract analogy with previous experience. This is in contrast to the child's capacity which is limited by the fact that the child is equipped with a smaller store of previous linguistic items from which to form analogies. Forming an analogy on the basis of fewer instances constitutes a process of production based on less abstract linguistic relations. It is interesting to note Gershkoff-Stowe's point that the immediate context of speech influences the nature of what is produced, but her alternative definition of abstract knowledge seems vague. I agree that non-linguistic domains are important, and indeed the UB approach is concerned also with establishing the consequences of linguistic usage for the attention required during speech production (e.g. directly accessible chunks of language are formed on the basis of frequent use and become automated, Vogel Sosa & MacFarlane, 2002), however I do not accept the position that the notion of permanent knowledge should be rejected entirely. It is sufficient, I believe, to acknowledge that permanent knowledge represents a less frequently used route to producing language, and that the immediate circumstances in which language is produced will impact which route the individual takes.

Skljarov

Skljarov (2004a) made two main points to which I would like to respond, and one more minor comment. The first main point concerns his rejection of what he terms the 'phonological approach' to understanding stuttering, and the second relates to his rejection of Savage and Lieven's (2004) claim that research has yet to resolve the issue of where the boundaries of planning units in speech production are located for children of various. The two issues will be approached in turn.

Skljarov (2004a) cited Throneburg, Yairi and Paden's (1994) study as evidence that there are problems with what he termed the 'phonological approach' to understanding childhood stuttering. This study showed that for CWS phonological difficulty was unrelated to incidence of stuttering, as measured by age of onset for consonants, syllable number in a word, and presence or absence of consonant strings. Similar results were found by Howell and Au Yeung (1995). Skljarov takes the view that this undermines the phonological approach, but seems not to take account of the later work by Howell and colleagues that showed an effect of phonological complexity on stuttering behavior when a more sensitive measure was used. Howell, Au Yeung and Sackin (2000) pointed out that neither the Throneburg et al. (1994) study nor that of Howell and Au Yeung (1995) took account of the position within a word that phonologically difficult items occurred, that is whether a late emerging consonant (LEC) or consonant string (CS) occurred at the beginning or end of a word, or if they occurred in the same place. This is particularly pertinent in light of the evidence that stuttering almost always occurs on the first phoneme of a word rather than at the end (Brown, 1945; Wingate, 1982, 1988). In fact, Howell, Au Yeung and Sackin (2000) found that the phonological factors of LEC and CS affected frequency of stuttering when word positions of these factors were considered, for children, teenagers and adults who stuttered. Speakers from all age groups stuttered significantly more on content words that started with CSs than those that contained no CS at all and the CWS stuttered significantly more on words starting with LECs.

Skljarov (2004a) argued that if phonological planning does not present a problem for CWS, then the system must break down at another level of speech processing, namely the rhythmic level (e.g. Skljarov, 2004b). Whilst I acknowledge that rhythmic planning may indeed be an important contributor to stuttering behaviour, I would disagree with his argument that the nature of the units involved across developmental time have been revealed from a UB perspective. Skljarov (2004a) seems to conclude that the rhythmical properties of speech can explain stuttering behaviour so thoroughly that no further explanatory power is to be drawn from other forms of planning, such as phonological or syntactic difficulty. I do not accept that the rhythmic model constitutes a 'catch-all' explanation. The evidence cited above from Howell and colleagues shows us that phonological difficulty is relevant when considered at the appropriate level of detail, but a UB model of how phonological planning units change over developmental time has yet to be developed. Likewise, as described in the Savage and Lieven article, it is clear that the units of syntactic representation that children use are different from those of adults, but we have yet to establish how these are used online. As noted by Gershkoff-Stowe (2004), there are other factors that also need to be taken into account (e.g. prosody, non-linguistic factors) and we must then establish how all these components interact. The speech planning process is very complex, and I maintain that there is a need to explore further the nature of children's planning units over time from the UB perspective.

Finally, there is a more minor point worth mentioning that concerns Skljarov's (2004a) interpretation of Gershkoff-Stowe's description of the language system of CWS moving from instability to stability as they grow older. Skljarov argues that this needs to be reversed to capture the nature of the system, to move from stability to instability. I believe that this disagreement may arise from different semantic interpretations of the terms stability and instability. Gershkoff-Stowe seems to be referring to a system wise stability that entails producing a stable end result. However, Skljarov

seems to refer instead to a contrast between inflexibility and flexibility or restricted versus dynamic. His argument seems to be that the system moves from an internal sort of 'stability', whereby the child can use only one form of rhythm, to a more 'instable' system that can produce various types. It would be necessary for the authors to further define their terms before we could establish whether this disagreement is based in theory or in semantics.

Schwarz

Martin Schwarz raises two interesting questions. First, he asks "*Since it is well known that the most typical form of the onset of stuttering is marked by "effortless" repetitions of sound, syllables or words at the beginning of sentences, would the authors care to comment precisely how their linguistic model would account for this exact behavior?*" This question is relatively straightforward to answer by combining the UB approach with the EXPLAN model (Howell & Au-Yeung, 2002) to explain the mechanism that lies behind incidences of stuttering in speech planned 'online' (in real time). The argument from the UB paper of Savage and Lieven (2004) is basically that a directly accessed automated 'chunk' of language, such as '*I dunno*' (identified by Bybee & Scheibman, 1999), removes the need for a speaker to perform a complex syntactic planning exercise during language production, whereas when a speaker generates a more creative (less frequently co-occurring) sentence, they are faced by the need to fill an abstract 'slot' in a usage based 'slot-and-frame' construction. This entails accessing a word that does not always fill that position relative to the other lexical items and placing it in position on the basis of its potential to fill that role in the construction only on an abstract basis, by analogy rather than frequently occurring concrete exemplars. The second process requires a heavier online planning load, such that the speaker's plan for the word may not be complete at the point when the 'frame' words have already been executed. It is likely that for young children at the onset of stuttering, the boundary of a 'slot-and-frame' schema will correlate with a traditional sentence boundary, because each sentence is only a few words long. Drawing on the EXPLAN model, we would then expect that the 'frame' word that precedes the 'slot' filling word would be repeated if the process of filling the slot was not complete. Whilst it is possible that a 'slot-and-frame' schema could end within a sentence when more than one is used to create an utterance, it is unlikely that the start of the schema will not also constitute the start of the sentence. This means that there would be enough coincidence of the schemas with sentences to give the picture of stuttering occurring at the beginnings of sentences near the age of onset.

Schwarz states that "*it has been documented that when children and adults who stutter speak to themselves out loud alone they are, in the main, totally fluent.*" He asks, "*Could the authors, using the model they have put forth, precisely explain this behavior?*" The variation in an individual's rate of stuttering is an important issue that does require explanation, though is not often directly considered by theories of stuttering. From the UB perspective adopted in the Savage and Lieven (2004) article, a return to fluency in some contexts could be explained for a number of reasons, which may vary between children and adults. The EXPLAN model (Howell & Au Yeung, 2002) would provide the hypothesis that fluency is attained when a speaker is able to allow sufficient time to plan a word before it is executed. This could stem from either being able to speak at a slow rate or from using more simple words. For either children or adults who stutter, it is plausible that their fluent speech when talking alone is a result of their allowing more planning time by slowing their speech rate because they are free from the time pressure that is generated by the context of joint turn-taking speech in engagement with an interlocutor. It is also possible that when alone, adults who stutter are free to use more easily accessible words or constructions (e.g. a frequent expression) because they are free from the conversational demands of creativity introduced by the need to respond to an interlocutor. It is an empirical question as to whether either of these variables are valid influences on either children or adults who stutter.

References

Bock, K. & Loebell, H. (1990). 'Framing sentences'. *Cognition*, **35**, 1-39.

Bock, K. & Griffin, Z. (2000). 'The persistence of structural priming: transient activation or implicit learning?'. *Journal of Experimental Psychology: General*, **129** (2), 177-192.

Bosshardt (2002). 'Effects of concurrent cognitive processing on the fluency of word repetition: comparison between persons who do and do not stutter'. *Journal of Fluency Disorders*, **27**, 93-114.

Bybee, J. (1998). 'The Emergent Lexicon'. *Papers from the Regional Meetings, Chicago Linguistic Society*, **2**, 421-435.

Bybee, J. & Scheibman, J. (1999). 'The effect of usage on degrees of constituency: the reduction of don't in English'. *Linguistics*, **37** (4), 575-596.

Brown, S. (1945). 'The loci of stutterings in the speech sequence'. *Journal of Speech Disorders*, **10**, 182–192.

Chang, F., Dell, G., Bock, K. & Griffin, Z. (2000). 'Structural priming as implicit learning: A comparison of models of sentence production'. *Journal of Psycholinguistic Research*, **29** (2), 217-229.

Gershkoff-Stowe, L. (2004). 'Commentary on Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?' *Stammering Research*, **1**, 101-102.

Howell, P. & Au-Yeung, J. (2002). 'The EXPLAN theory of fluency control and the diagnosis of stuttering'. In E. Fava (ed.), *Current Issues in Linguistic Theory series: Pathology and therapy of speech disorders* (pp.75-94). Amsterdam: John Benjamins.

Howell, P., & Au-Yeung, J. (1995). 'The association between stuttering, Brown's factors and phonological categories in child stutterers ranging in age between 2 and 12 years'. *Journal of Fluency Disorders*, **20**, 331–344.

Howell, P., Au-Yeung, J. & Sackin, S. (2000). 'Internal structure of content words leading to lifespan differences in phonological difficulty in stuttering'. *Journal of Fluency Disorders*, **25**, 1-20.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Bradford Books.

Pickering, M. & Branigan, H. (1999). 'Syntactic priming in language production'. *Trends in Cognitive Science*, **3** (4), 136-141.

Savage, C., & Lieven, E. (2004). Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering? *Stammering Research*, **1**, 83-100.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (under review for *Language Learning and Development*). 'Structural priming as implicit learning in language acquisition: The persistence of lexical and structural priming in 4-year-olds'.

Schwarz, M. (2004). 'Commentary on Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?' *Stammering Research*, **1**, 106.

Skljarov, O. (2004a). 'Commentary on Can the Usage-Based Approach to Language Development be Applied to Analysis of Developmental Stuttering?'. *Stammering Research*, **1**, 103-105.

Skljarov, O. (2004b). 'Speech development and scenario of its V-rhythms' development'. *Technical Acoustics*, **7** (Electronic Journal: http:/webcenter.ru/~eeaa/ejta).

Throneburg, N., Yairi, E. & Paden, E. (1994). 'The relation between phonological difficulty and the occurrence of disfluencies in the early stage of stuttering'. *Journal of Speech and Hearing Research*, **37**, 504–509.

Vogel Sosa, A. & MacFarlane, J. (2002). 'Evidence for frequency-based constituents in the mental lexicon: collocations involving the word *of*'. *Brain and Language*, **83**, 227-236.

Wingate, M. (1982). 'Early position and stuttering occurrence'. *Journal of Fluency Disorders*, **7***,* 243-258.

Wingate, M. (1988). *The structure of stuttering: A psycholinguistic analysis.Berlin:* Springer-Verlag.

TARGET ARTICLE
Involvement of social factors in stuttering: A review and assessment of current methodology

Adrian Furnham and Stephen Davis
Department of Psychology, University College London, Gower Street, London WC1E 6BT
a.furnham@ucl.ac.uk
stephen.davis@ucl.ac.uk

Abstract. Most models that explain the onset and development of stuttering include a social and emotional component. This paper has two intentions. One is to review the methods and findings of previous research that investigated the role of affective and social factors in stuttering. The second intention is to alert readers to various methods and issues being applied in social psychology to investigate these phenomena and to indicate where these methods could be useful in assessing the role of social and affective components in stuttering.
Keywords: Intelligence, personality attitudes, temperament, bullying, self esteem and stigma, anxiety, occupation.

1. Background

Stuttering affects approximately 5% of the population at some time in their life according to work in the United States (Conture, 1996). It is also reported that the disorder affects children disproportionately. The usual age at which the problem starts (onset) is between three and five years (Dalton & Hardcastle, 1977). Eighty per cent of young children who are diagnosed as stuttering recover to normal fluency during school years (Starkweather, 1985; Yairi & Ambrose, 1999). Around one in a hundred of the adult population persist in their stuttering (Andrews & Harris, 1964; Bloodstein, 1987). It is important to know what role social factors, and their changing role over development, play in stuttering. This is because speech is a social phenomenon as people speak to one another about various topics in a variety of situations. Also, stuttering has been shown to be topic- and situation-specific and the disorder is governed in part by affective factors that are socially moderated.

The disorder also affects language and motor performance. Historically, the evidence that showed deficits in these areas of performance led to accounts which maintained that stuttering arises either because of psychological, physiological, linguistic or learned behaviors. In contrast to such views, accounts have emerged over the last quarter of a century, that maintain stuttering is a multifaceted speech disorder that involves all the preceding factors and, in addition, social ones. There have been several models that approach stuttering from a multifactor perspective, most of which include a social or emotional component. One of the earliest multifactor models of stuttering (Zimmerman, 1980) emphasised the importance of the interaction between motor speech behavior and a range of emotional and environmental conditions in the development and maintenance of the disorder. Wall and Myers (1984) suggested that psycholinguistic factors, psychosocial factors (i.e. discourse loads and interactions with parents and/or peers) and physiological components interacted to cause and maintain stuttering. The demands and capacities model (Starkweather, Gottwald & Halfond, 1990) views the onset and development of stuttering as related to a mismatch between the child's capacities (motor, linguistic, cognitive and emotional) and self-imposed or external speech demands. The models proposed by Smith (1999) and de Nil (1999) concentrate on the importance of disrupted speech processes and their relationship with social, emotional and learned factors. All these models suggest that cognitive, linguistic and affective factors influence speech motor functions. Riley and Riley (2000) revised their 1979 stuttering assessment instrument and maintained that speaker temperament factors and listeners reactions to people who stutter were two of the three main factors that contributed to the onset of stuttering.

Multifactor models describe how stuttering might start and/or be maintained and, by implication, indicate that components identified in the models should be included in the identification and treatment of the disorder. However, with the possible exception of the Riley and Riley work, the models lack the detail necessary to make them useful in the collection and interpretation of assessment and treatment outcome data (Healy, Trautman & Susca, 2004). The influence of social and emotional factors in the onset, development and treatment of stuttering is ubiquitous in the models. The case for inclusion of social aspects in stuttering is aided further by confirmation of influences of social variables on stuttering (in both onset and treatment) using the most rigorous techniques applied in this area.

This paper has two intentions. One is to review the methods and findings of previous research that has investigated the role of affective and social factors in stuttering. This will constitute an examination

of the research that investigated differences between people who stutter, PWS (children and adults) and people who do not stutter and will also look at research investigating how others perceive PWS. This will include studies of how PWS (children and adults) are perceived by fluent children and adults (including clinicians and parents). A common finding of past research is that PWS are stereotyped as being more guarded, nervous, self-conscious, tense, sensitive, hesitant, introverted, and insecure than speakers who do not stutter (Klassen, 2001). In cases where there is an influence, they can be included in multifactor models that address the onset, development and treatment of the disorder. The second intention is to alert readers to various methods and issues being applied in social psychology to investigate these phenomena and to indicate where these methods could be useful in assessing the role of social and affective components in stuttering.

Social factors are examined around the start of school, at adolescence and in adulthood. These age groups are looked at separately as different social factors kick in as speakers get older. Thus the factors that will be considered with pre-school and start of school children are intelligence, personality, attitudes and temperament, bullying can occur once a child is established in a school and progress through adolescence and probably into adulthood and occupational matters specifically affect adults. Factors affecting one of the early age groups can have an influence through life and such factors have sometimes been investigated in later age groups. Thus, though intelligence exerts an effect on young CWS, the question can also be raised whether it also affects adolescents or adults. Factors are considered in the age groups where they may first exert an influence, but evidence on the same factor in older age groups is also considered so developmental trends can be established. The target group at the first two ages is referred to as children who stutter (CWS) and the older age group as persons who stutter (PWS). All PWS is used to refer to all ages. Children who do not stutter (CWNS) and adults who do not stutter (PWNS) represent controls. CWS have been evaluated by peers (CWNS) and by adults (PWS or PWNS). Adult PWS have also been quizzed retrospectively about how social influences affected their stuttering in their childhood. An attempt has been made to distinguish these different options in the studies reviewed.

2. Pre school and start of school

Differential psychologists distinguish between two major variables: 'Abilities' measured by tests of statistical power (how robust the statistical findings are) and 'personality' measured by tests of preference. The latter is usually further sub-divided into beliefs (attitudes, attributions, values) traits (types, disorders) and coping styles. To get a full understanding of the role of individual differences in stuttering, it is important to understand all aspects of individual differences that may act as moderator variables on stuttering.

Intelligence

In the intelligence researcher community the world is divided into 'lumpers' and 'splitters'. The former believes in general, the latter in multiple, intelligence. All intelligence tests, particularly those of crystallized intelligence, have verbal tests (usually vocabulary tests). Indeed, vocabulary is very highly loaded on 'g' that measures general intelligence (REF). It is, therefore, a valid hypothesis that a person with low IQ manifests in part by low vocabulary may be particularly frustrated or anxious in social settings when they are required to articulate.

Empirical evidence suggests that CWS score significantly lower on intelligence tests than do fluent controls. Studies with school-age CWS (Andrews & Harris, 1964, Okasha, Bishry, Kamel & Hassan, 1974, Schindler, 1955) indicated that this deficit is evident in both verbal and non-verbal intelligence tests. As non-verbal intelligence is affected, it appears unlikely that these performance deficits could be explained by difficulties in communication because of stuttering. In contrast, the intelligence and social class of CWS who are receiving treatment is found to be above average (Andrews & Harris, 1964, Cox, 1982). Though a developmental change could be responsible, the difference between younger and older CWS could also be explained by the influence of intelligence and social class on access to health care.

The intelligence of all PWS as perceived by others appears to change across age groups. Franck, Jackson, Pimentel and Greenwood (2003) used bi-polar adjective pairs taken from Freeby and Madison (1989) and Wencker et al (1996). Examples of adjective pairs used were *intelligent-stupid* and *competent-incompetent*. Fluent children aged 9 - 11 years were asked to view videotapes of an adult who stuttered and a fluent control. They found that the CWNS used adjectives that indicated they judged the intelligence of the PWS more negatively than the PWNS. However Craig, Tran and Craig (2003) reported that telephone interviews with 502 adults indicated that a large number of the respondents (none of whom stuttered or had any interaction with CWS or PWS) believed that PWS had average or above average intelligence. Taking the results together, there seems to be a difference between the way children perceive PWS (less intelligent) and the way adults perceive PWS (more intelligent).

Personality

Personality traits have been shown to relate to second language learning and work productivity (REFS). Indeed they have been shown to account for a tenth to a third of the variance on a wide range of factors like academic success, work productivity and health. It seems quite plausible that traits, particularly extraversion and neuroticism, both as main effects and in interaction, relate to speech fluency and stuttering. Indeed fluency versus disfluency may affect, in turn, the personality of young people. What, however, is important is a description of the process by which stable traits relate to problems in language production.

Early descriptions of stuttering regarded it as a manifestation of emotional disorder in childhood. Considerable effort was invested in collecting data on the personality attributes of PWS and their tendency to show anxiety or neurotic symptoms that would support such descriptions. The work used a variety of measurement instruments, and indicated that there were no differences between school-age children who stutter and controls in personality factors related to neuroticism or anxiety. This conclusion is based on tests with the Sarson General Anxiety Scale for Children and the Structured Psychiatric Interview (Andrews & Harris, 1964); Eysenck Personality Inventory (Hegde, 1972); California Test of Personality (Prins, 1972; Minnesota Multiphasic Personality Inventory (Horlick & Miller, 1960, Lanyon, Goldsworty and Lanyon, 1978, Pizzat, 1951); and the Speolberger Anxiety Scales (Molt & Gifford, 1979).

Manning, Dailey and Wallace (1984) found that self-perceived personality characteristics of older PWS (29 participants, mean age 62 years) were not significantly different from those of nonstuttering controls (13 participants, mean age 65 years). They used a bi-polar adjective scale (Wood & Williams, 1976) containing pairs such as *anxious-composed* and *introverted-extroverted*. The PWS had a tendency to see themselves as more inflexible, withdrawn, self conscious, anxious and introverted than the control group, however no significant differences between the groups were reported.

Personality alone does not appear to be a predictor of the developmental pathway of stuttering. Guitar (1976) showed that neither neuroticism or extraversion as measured by the Eysenck Personality Inventory (Eysenck & Eysenck, 1963) were, by themselves, significant predictors of recovery or persistence. However Guitar did conclude that a combination of pre-treatment factors (e.g. personality traits, % stuttered syllables, attitudes) was useful in predicting the outcome of speech therapy. His study indicated that a combination of measure of several pre-treatment factors taken on 20 adults who stutter was highly correlated with post-treatment speech measures (% stuttered syllables, % change in frequency of stuttering). Guitar (1976) concluded that, of the personality traits, only neuroticism was strongly related to outcome measures. However, in this study neuroticism was also found to significantly relate to attitude measures.

Other studies have looked at the way the personality of PWS is perceived (as opposed to measuring what it is actually like). For instance, PWS are generally stereotyped as more nervous, shy, withdrawn, tense and anxious (Horsley & FitzGibbon, 1987). The Franck et al (2003) study mentioned above used the adjective pairs from Freeby & Madison (1989) and Wencker et al (1996) and included pairs such as *outgoing-shy* and *relaxed-tense*. They found that fluent school-age children rated adults who stutter more negatively than controls on a series of personality traits. However, unlike the rating of the intelligence of PWS, the negative perception of the personality of PWS continues into adulthood. Such stereotypes have been found to be held by a wide range of groups, from members of the general public, including college professors and teachers and even clinicians (Dorsey & Guenther 2000; St Louis & Lass, 1981; Yeakle & Cooper, 1986) as well as the parents of CWS (Fowlie & Cooper, 1978).

Wood and Williams (1976) found evidence for a strong negative stereotype of CWS. Using an analysis of responses to bi-polar scales derived from words previously judged by speech clinicians to be descriptive of CWS (e.g. *nervous-calm, afraid-confident*) they found the stereotype to be predominantly unfavorable. Wood and Williams (1971) found that even people with professional experience of working with disfluent patients attributed undesirable characteristics to CWS. When asked to list adjectives to describe CWS, approximately 75% of clinicians used words that were grouped within the category of "nervous and fearful" and 64% listed words that were included in the category "shy and insecure".

Looked at from the broader social psychology perspective, De Waele and Furnham (1999), in a meta-analysis of the role of trait-extroversion, found that linguists were very out-of-date in their understanding and measurement of personality traits. Also they were seriously misinformed about the role of personality in language. De Waele and Furnham's own analysis showed clear and consistent evidence of the role of extraversion in language learning and production. Another important limitation in many studies is that many researchers examining stuttering have not looked at interactions. The interaction of extraversion, neuroticism and intelligence may well show powerful effects where looking at these factors alone does not.

The best way to determine the role of personality variables would be incrementally through block stepwise regression. Thus, assuming one has a sensitive and robust criterion variable like degree of stuttering or some measure of recovery, one could regress demographic variables first (age, sex, class), then intelligence, and finally personality measures using the big five personality dimensions (as main and interaction effects). This would then demonstrate if and whether personality traits accounted for any incremental variance in explaining stuttering-related variables.

Attitudes

The attitudes and reactions of all PWS to interpersonal verbal communication (communication attitudes) have been regarded as constituting a basic component of stuttering for many years (van Riper, 1948; Johnson, Brown, Curtis, Edney & Keaster, 1956; Travis, 1957; Sheehan, 1970; De Nil & Brutten, 1991; Vanryckegham & Brutten, 1996). Several studies have produced evidence that the communication attitudes of adult PWS are more negative than those of adult PWNS (Brown & Hull, 1942; Erikson, 1969, Andrews & Cutler, 1974).

According to Costello (1984), few attempts have been made to assess the attitudes of CWS. Two reasons suggest that measuring attitude in pre-school children is important: i) The most important point that applies to speakers of all ages is that attitudes have been shown (in particular circumstances) to be causally-related to behavior. ii) Stuttering usually starts between three and five years (Dalton & Hardcastle, 1977) Attitudes about many things change dramatically between ages 3 and 5 (Perry, Bussey & Fischer, 1980). This is especially likely to be the case when a child has a problem that affects overt behavior. An observation which may be relevant to effective treatment is that attitudes are more easily changed during or close to their formation (Niven, 1994). Though all these points underline the importance of having procedures and measured of attitude in young people who stutter, methodologically the development of appropriate instruments is not straightforward.

One attempt was made by Brutten (1985) who developed the Communication Attitude Test (CAT) in order to determine if the speech-related attitudes of CWS differed from those of CWNS. A Dutch version of the CAT (CAT-D) was developed in a series of studies that aimed to establish if the communication attitudes of CWS were significantly different to those of CWNS for speakers of this language (De Nil & Brutten, 1986; 1991). These studies revealed that CWS scored significantly higher on the CAT-D than their peers who did not stutter (i.e. CWNS), indicating that their speech-related attitudes were more negative. Similar between group differences were found with a group of American children by Boutsen and Brutten, (1989). The internal, and test-retest, reliability of the CAT and CAT-D has been demonstrated in several studies (Brutten & Dunham, 1989; Vanryckeghem & Brutten, 1992; Vanryckeghem & Brutten, 1992).

Vanryckeghem (1995) proposed that the CAT and CAT-D are useful clinical and research tools for evaluating between group differences when investigating communication attitudes. However she considered that the scope of the CAT is limited in that it requires a child to have the ability to read and understand the concepts covered by the test items and consequently it is not generally accurate when used with children younger than 7 years of age.

An instrument capable of determining the communication attitudes of children close to stuttering onset would be useful in several areas. In research it would assist in pinpointing the exact role of communication attitudes in the onset and development of stuttering. There have been proposals that the onset of stuttering is a result of the belief that speech is difficult (Bloodstein, 1987; Brutten & Dunham, 1989). Diametrically opposed to this viewpoint, are theories that have proposed that the negative beliefs that PWS have about speech are a product of, rather than a cause of their dysfluency (Guitar, 1976; Peters & Guitar, 1991). In terms of clinical efficacy Shearer (1961) and Erikson (1969) indicated that changes in the self-concept of someone who stutters is an important aspect of success both during and following treatment. An instrument that offers an indication of attitude in young children might also be useful for early diagnosis and intervention (Yairi & Ambrose, 1992: Onslow, 1994).

Attitudes to stuttering are comparatively rare in older speakers but there are at least four areas of research into attitudes that merit further work in older (as well as younger populations). First, there are the attitudes of speakers who stutter to themselves and to other speakers who stutter in social situations in which they stutter. Second, the attitudee of speakers who stutter to social communication in general need examining as they might yield patterns of situational phobia or anxiety. Third, there are the attitudes of various groups – parents, teachers, peers and the public – to those with different patterns and degrees of stuttering. Fourth, and perhaps most importantly, models like the theory of planned action need to be employed to determine how, when and why attitude in observers relate to their behavior towards them. Ideally this would involve costly, but ever important, longitudinal research.

Temperament

Many authors, such as Sermas and Cox (1982), hold that a child's temperament is a contributing factor to both the development and maintenance of stuttering. Temperament is distinguished from traits being much more physiologically based. Theorists in the last decade or so have suggested that CWS

exhibit a more vulnerable or sensitive temperament, which could possibly be a contributing factor in the development, maintenance or chances in recovery of stuttering (e.g., Conture, 2001; Guitar, 1998; Zebrowski & Conture, 1998). As well as temperament as a whole, it has been suggested that particular aspects or dimensions encompassed in temperament play a role in stuttering, for instance attending problems, which refer to distractibility, perseveration, inability to concentrate on tasks and low frustration tolerance (Riley & Riley, 2000). In the few studies that have addressed stuttering and temperament directly, a number of dimensions have been highlighted as differing in CWS relative to CWNS. In general, CWS tend to be more responsive or reactive to stimuli in their environment (Wakaba, 1998), and are more sensitive, anxious, withdrawn and introverted (Fowlie & Cooper, 1978). These findings support the speculations of Bloodstein (1995) and Guitar (1998) that CWS have a more sensitive temperament leading to greater reactivity to unfamiliar, challenging or threatening situations, thus supporting the notion of multiple dimensions of temperament.

Embrechts, Ebben, Franke and van de Poel (2000) assessed temperamental dimensions using parental reports on the Children's Behaviour Questionnaire (Rothbart & Bates, 1998), and found that parents rated their child who stutters as displaying reduced attention span and less success in adapting to new environments. More recently, Anderson, Pellowski, Conture & Kelly (2003) found, using the Children's Behaviour Questionnaire, that CWS were significantly less likely to adapt to change, were less distractible and displayed greater irregularity with biological functions.

However, there are discrepancies in findings in terms of what dimensions CWS differ from CWNS. In the Embrechts et al. (2000) and Anderson et al. (2003) papers, two dimensions were found to be significantly different between CWS and CWNS: First, CWS were found to be significantly less adaptable that their controls. A second difference is in terms of distractibility. Here, however, though this dimension was found to be significantly different in both the studies cited, the direction was not the same; in Embrechts et al., CWS were more distractible than the controls, but in Anderson et al., they were found to be less distractible. In addition to these two dimensions, Anderson et al. found that CWS were significantly less rhythmic than CWNS in terms of the rhythmicity dimension, whereas Embrechts et al. did not. In a paper by Howell, Davis, Patel, Cuniffe, Downing-Wilson, Au-Yeung and Williams (in press), four dimensions were found to differ significantly. As was the case for both Anderson et al. and Embrechts et al. studies, CWS were found to be non-adaptable. However, none of the remaining three dimensions matched in direction across this and the two cited studies; Howell et al. (in press) found, in particular, that CWS were significantly more active, more negative in mood, and less persistent than CWNS.

Future work should examine Eastern European models of temperament, it is possible to test both cross-sectionally and longitudinally whether temperament variables relate to stuttering. Temperament variables, like 'strength of the nervous system' might lead to problems in speech production, particularly in stressful situations. Temperament factors might link, then, to physiological predispositions to stuttering. Furthermore, temperamental factors might indicate the likely success or otherwise of effective treatments (assuming disorders with physiological underpinnings are less likely to be treated effectively). Surprisingly no research as yet has looked at whether temperament predicts the onset and likelihood of recovery in children who stutter.

3. School and adolescence

Bullying

Two fundamental questions are 1) whether speakers who stutter are bullied and 2) whether bullying exacerbates the problem of stuttering? There only appears to be evidence on the former.

Parker and Asher (1987) reviewed work on bullying and concluded peer rejection and bullying can have severe and long-lasting effects such as low peer acceptance or peer rejection and their influence on later personal adjustment problems such as depression and early school dropout. Hodges and Parry (1996) identified three peer-related factors that increased the risk of a child being bullied - few friends, low-status friends and rejection by peers. O'Moore and Hillery (1989), Martlew and Hodson (1991), Nabuzoka and Smith (1993) and Whitney, Smith and Thompson (1994) have all reported that children with special educational needs are more susceptible to bullying than their peers, and are more likely to have few friends and be rejected.

The reason for thinking social acceptance might affect children who stutter is that they are often reluctant or unable to participate verbally in school activities (or social groups in general). In turn, this may lead them to be seen as shy or withdrawn and possibly, because of these perceived characteristics, to have difficulties in peer relationships making them targets of bullying. There is some previous research investigating sociodynamic factors and their relationship to stuttering and this is arranged for review under the techniques they have used (ratings, retrospective, sociometric). It is apparent that the majority of this research was carried out over a quarter of a century ago and, although useful from an

historical perspective, may or may not be relevant to present practice (thus, highlighting the need for work on this topic in current schools).

1. Ratings by fluent adults of children who stutter. Perrin (1954) found that CWS were not readily accepted as members of their classroom, and suggested that stuttering students who were more able to adjust to interpersonal situations responded better to therapy. Though some studies present a coherent picture of problems a CWS has in different social groups, other studies have failed to find any influence of social factors on the status of CWS. Brissey and Trotter (1955) examined the social relationships within a group of speech-impaired children enrolled in a six-week summer residential clinic. They found no indication that social status was correlated with the severity of speech impediment. They concluded that the composition of the study group may have led the members to be more tolerant of speech dysfluencies compared with when one stuttering child had been in a group of fluent peers. Woods (1974) reported evidence that showed the social relationships of a stuttering child were no better or worse than those of fluent classmates. These studies using early clinical investigative tools, besides being dated, have not resulted consistent findings.

Many other groups can do observational data: parents, peers, teachers being obvious examples. Observational data have the advantage of rater-reliability in the sense that different raters can be compared as this can serve as an index of reliability. However, raters see their 'target' in different contexts that may indeed affect the rating. Thus teachers see classroom behavior, parents dining table behavior and peers contact with strangers. Observers have qualitatively and quantitatively different data. Thus low reliability (poor alphas) may not be an index of poor ratings.

2. Retrospective self-ratings by adults who stutter about bullying in childhood. Retrospective reports of problems in social groups by PWS have also been obtained. Mooney and Smith (1995) used a questionnaire to obtain information regarding their time at school from adults who stutter. They found 11% of adults who stutter said they had been bullied at school and that this had a negative effect on the fluency of their speech. Comparison of this with estimates about how many fluent school children are bullied, on the other hand, indicates that children who stutter are no more at risk of bullying than their peers. Haynie, Nansel, Eitel, Crump, Saylor, Yu and Simons-Morton, (2001) reported that over 30% of schoolchildren stated that they had been bullied within the last school year (much higher than the 11% of adults that stutter who reported having been bullied by Mooney and Smith, 1994).

The most recent retrospective report was by Hugh-Jones and Smith (1999). In this study 74% of 276 adults who stutter that took part in the survey reported that they had been bullied during their time at school. Of the 205 respondents that indicated they were bullied at school, 6% reported that the bullying had a long-term effect on their fluency. However, the study lacked a control group who does not stutter to establish whether fluent speakers were bullied less often. The authors also note other limitations in the project, common to all retrospective studies: Respondent's recollections may be distorted and there is no way of validating the responses. The authors also concede that the sample may be limited by the fact that the respondents were a volunteer sample from the British Stammering Association that may have resulted in a cohort that was particularly aware of the issues surrounding their dysfluency and its effects.

The usefulness of data from retrospective studies is limited by the contradictory findings. Also, retrospective self-reports are filled with methodological problems. Besides the problems already mentioned, studies that require retrospective reports (such as on parenting styles) are regularly rejected by journals as authors want to assert causality in that these styles affect adult behavior patterns. The reason for rejection is that these reports are compromised by the variable being investigated. Thus, for instance, it is possible that those who successfully 'recover' from stuttering offer a very different set of descriptives/ explanations than those who do not. These attributions and selective memory problems have led many to seriously question such data.

3. Sociometric assessment of fluent children rating children who stutter and children who stutter rating fluent children. Sociometric methods have also been used to assess the dynamics of groups containing CWS. Marge (1966) reported a study that used these procedures to assess intellectual and social status, physical ability and speech skills of CWS. The study examined 197 third grade (8-9 years) public school students, of whom 36 had been diagnosed with moderate or severe speech dysfluency. Sociograms were obtained on each of the four components investigated based on Moreno (1960). The study required children to rate other children in the class by responding to statements such as "I would like to work with this child" or "I would like to play with this child". Marge reported that dysfluent children held a lower social position than fluent ones. Sociograms were also obtained from teachers on the same four component skills for each child. The data from the study indicated that, with regard to intellectual skills in school and social activity outside school, the child who stuttered held a significantly lower position than that of his or her fluent peers. In the other areas of playground activity and speech skills, no significant differences between the groups were found. The

results from the teachers corroborated the findings of the peers. More recently Davis, Howell and Cook (2002) used a forced choice sociometric scale to assess the peer relationships of 16 CWS and their 403 classmates. They reported that CWS were rejected significantly more often than their fluent peers, were categorized as less popular and were less likely to be named as leaders. The CWS were three times more likely to be identified as victims of bullying than their fluent peers and generally demonstrated low social acceptance among peers.

A general problem to note that cuts across assessment methods used in studies of bullying, is that the majority of research into the social status of children who stutter uses data from respondents who were in the educational system more than two decades ago (either because the publications are dated or adult respondents provided retrospective reports). It is possible that the attitude of children toward their peers with disabilities (including those of speech) has changed in the intervening period. A secons general point is that researchers have always wisely called for the triangulation of methodologies to help reduce biases associated with each. There is no reason to believe that research in stuttering should be any different. Finally, it is possible to use other indexes of peer relationships. Diaries can prove useful as they can indicate the social world of the person who stutters. Phones can be logged to access number and length of calls.

Self esteem & stigma

There is consensus among clinicians and researchers working with PWS that speech disorders can have detrimental effects on self-perception and, specifically, on self-esteem (Bajina; Luper; Shames; Starkweather & Van Riper). As a result, therapeutic interventions often include either implicit or explicit goals to improve an individual's concept of self-worth (Bloodstein; Cooper; LaBlance; Luper &). Yet, there are minimal empirical data that indicate a need for the implementation of regular clinical attention to self esteem for PWS in general (Yovetich, Leschied & Flicht, 2000).

Self-perception and self-concept have often been addressed in therapy for PWS (Sheehan; Silverman & Van Riper). Sheehan and Martyn (1966) suggest that individuals who have developed a concept of self as a PWS are less likely to recover spontaneously than those who have not. Beach and Fransella (1968), however, believe that for therapy to be successful, PWS must accept their speech disorder as part of their self-concept.

Children and adolescents who stutter self-rating of self esteem. Pukacova (1973) used a projective technique (incomplete sentences) to estimate the self-esteem of 74 CWS; 94% of this sample evidenced low self-esteem. More recent work (Yovetich et al 2000; Blood, Blood, Tellis & Gabel, 2003) found no evidence for low self-esteem for school-age and adolescent stutterers, Yovetich et al (2000) used Battle's (1992) Culture Free Self-Esteem Inventory. They reported no differences between the mean scores for CWS and the normative data. Eighty per cent of the participants (school age CWS) scored above the standardised mean on the Total Self-Esteem score. Blood et al (2003) found that 85% of the adolescents who stutter that participated in the study scored within 1 SD from the mean on a standardised measure of self-esteem – the Rosenberg Self-Esteem Scale (Rosenberg, 1965) meaning they were not significantly different to the norms.

Adults who stutter self-rating of self esteem. Bardrick and Sheehan (1956) found that individuals with lower self-esteem showed higher rates of stuttering. Bajina (1995) noted a similar trend toward lower self-esteem in 28 PWS. Shames and Rubin (1986) report that the most common attitudes expressed by PWS are anxiety, helplessness, victimization, and low self-esteem.

To summarize the work on self esteem, how speakers who stutter perceive themselves may be decided in terms of their personality, their demographic background and also the severity of their stuttering. Though this topic appears to be 'easy' to research, it is hard to do well. The reason for this is threefold. First, a good control group is needed and a large representative sample for control speakers and speakers who stutter. Next, it is important to measure and analyze all the factors that relate to self esteem. This is because stuttering may be a moderator or mediator variable rather than a variable that affects stuttering directly. Most important, it is desirable to do longitudinal research to determine whether, why and when self esteem is a cause, consequence or both of stuttering.

Anxiety

Absence of data on anxiety levels of CWS? Anxiety is widely believed to a causal factor in stuttering and plays a central role in many theories about the origin of stuttering (Miller & Watson, 1992). Stuttering usually worsens when a PWS speaks to strangers or addresses large audiences or those felt to be his or her superiors. This often leads to PWS avoiding social and public speaking situations and experiencing anxiety when in those situations (Van Riper, 1992). Yet PWS may not differ from those who do not stutter in baseline levels of anxiety (Blood & Miller), although this finding has been questioned (Craig, 1992).

Preliminary searches indicate that there has been no empirical work conducted examining the anxiety levels of children who stutter when compared to fluent controls. Perhaps this is an oversight

and, if so, hopefully this will be highlighted in commentaries. The lack of data from CWS does not allow for causal implications to be considered. Is stuttering a consequence of anxiety or are adults who stutter more anxious because they stutter?

Self-reports of anxiety by adults who stutter. Several methods have shown consistently that anxiety levels in adult PWS are higher than those of adults who do not stutter. Trait anxiety questionnaire (Craig, Hancock, Tran & Craig, 2003); Inventory of Interpersonal Situations (IIS) (Kraaimaat, Vanryckeghem & Van Dam-Baggen, 2002); and the Cognitive Anxiety Scale (DiLollo, Manning & Neimeyer, 2003). Stein, Baird and Walker, (1996) found that seven out of 16 adults seeking treatment for stuttering could be classified as social phobics when using the DSM-IV criteria. Blood, Blood, Bennett, Simpson and Susman, (1994) found no differences between the anxiety levels of adult PWS who stutter and those of fluent controls using the State Anxiety Inventory, Trait Anxiety Inventory or Personal Report of Communication Apprehension. However, they did find that during high stress situations levels of salivary cortisol was significantly greater in the adult PWS than in the control group, indicating higher anxiety levels for the adults who stutter.

Miller and Watson (1992) examined self-perceptions of state and trait anxiety and refuted the assertion that PWS are more anxious than those who do not. Anxiety was not related to stuttering severity and their results indicated that high levels of anxiety in adult PWS were restricted to communications situations.

Anxiety can be measured by self-report measures of trait or state anxiety, observational reports of others or physiological measures like heart beat, galvanic skin response etc. In most research, there are surprisingly low correlations between them. For some it is subjective anxiety that is crucial and they advocate self report. Other researchers prefer physiological measures irrespective of whether speakers who stutter claim to feel stressed or merely aroused.

4. Adulthood

Occupational

There is some evidence that negative stereotypes and perceptions can spread and lead to role entrapment, or limited career choices, for women, minorities, people with physical disabilities, and individuals with hearing impairment with some interesting high celebrity exceptions. In the light of this conclusion, it is important to explore whether the negative stereotypes of stuttering will spread and lead to limited career choices for PWS. Gabel, Blood, Tellis and Althouse (2004) concluded that to date, no research has explored whether this applies. However, two types of research studies have explored related issues.

Adults who stutter self-reports of employment problems. The first type of research in this area involves the experiences of PWS at work. In a study exploring these issues, Rice and Kroll (1997) surveyed 568 National Stuttering Project members regarding their perceptions of past work experiences. Results indicated that stuttering directly affected these individuals' perceptions of work experiences and career choices. In particular, 70% of the participants reported that they believed they could have had a better job if they did not stutter, and 56% reported choosing a career that required less speaking. 35% of the participants reported that they believed stuttering had affected their chances of being promoted, reported feeling discriminated against in the hiring process, and perceived that their supervisors had misjudged their performance because they stuttered. In a similar study, Opp, Hayden, and Cottrell (1997) surveyed 166 PWS. These participants answered questions about their job choices, number of years they were employed, and whether or not they reported experiencing discrimination in their careers. Results of the study suggested that 35% of the participants reported being in careers that required a low level of communication, and 39% believed they had experienced discrimination in the hiring process because of their stuttering.

To summarize, these studies have found that a significant number of people who stutter believed that: (a) they could have had a better job if they did not stutter; (b) they chose careers that required less communication; and (c) they felt discriminated against in the hiring process. Neither of these studies offered any explanations about why PWS reported these experiences. These perceptions may have occurred because of the participant's own insecurities and beliefs related to stuttering or the negative perceptions of others in their environment.

It should also be noted that the Rice and Kroll (1997) and Opp, et al (1997) studies reviewed above suffer from the ubiquitous methodological problems allied to self-report data that have been addressed elsewhere in this paper. Briefly, there is no way of validating the respondent's reports and those responses may also be distorted by the passage of time. These attributions and selective memory problems have led many to seriously question this type of data. Although the nature of both studies make it difficult to include a control group, the data may be further compromised by the fact that the

respondents were a volunteer sample. This may have produced a cohort that was particularly aware of the issues surrounding stuttering and employment.

Employers perceptions about PWS who seek treatment and those who do not. Craig and Calver (1991) xplored issues related to employment of PWS who had attended a treatment program. The first part of the study surveyed 34 employers regarding their perceptions of two groups of PWS who were employed at their companies. One group of individuals who stuttered received therapy to speak more fluently, while the other group did not. The employers perceived the speech of the individuals who had completed treatment to be more acceptable and the perceptions toward the individuals who had not received therapy did not change. Results indicated that PWS were perceived more positively when they are able to improve their fluency. In addition, Craig and Calver (1991) polled the 62 individuals who had completed therapy program regarding vocation and career changes following treatment. Nineteen of the respondents reported a promotion following the completion of therapy and 18 responded that a positive job change (an upgrade from their former position) followed treatment. Results of this study suggested that people who stuttered were not only perceived in a more positive manner by employers following treatment, but also experienced a positive change in career. Scloss, Espin, Smith and Suffolk (1987) also found that employers produced significantly more favorable ratings of PWS at employment interviews after they had received specialist assertiveness training relating to their stutter.

Fluent adults rating adults who stutter at work. There is a third perspective – how employers and others view employment prospects for PWS. Gabel et al (2004) asked university students to report their perceptions of appropriate career choices for PWS. Results suggested that university students reported a perception that stuttering affected career opportunities and that 20 careers were judged to be unsuitable for PWS. Conversely, 23 careers were judged to be more appropriate for PWS than for fluent controls. In another study (Silverman & Paynter, 1990) students rated four scenarios ("a lawyer", a lawyer who stutters", a factory worker", and "a factory worker who stutters"). Both the lawyer and factory worker who stuttered were rated as being less competent than others in these occupations. The negative impact on the "appearance of competence" was greater for the former than for latter. Craig and Calver (1991) compared employer perceptions of their employees' speech between a group who had received treatment for stuttering and a nontreatment control. The employers' perceptions of the treatment group were significantly enhanced, whereas no significant change occurred in employers' perceptions for the control group. Hurst & Cooper (1983) used a questionnaire that required employers to indicate their strength of agreement to seven attitudinal statements concerning stuttering. Employers rejected the suggestion that stuttering has a negative effect on job performance but agreed that stuttering decreases employment prospects and interferes with promotion opportunities.

As a cautionary footnote, people in the business of recruitment and selection are sensitive to issues around discrimination. Unless the job specifically concerns/ necessitates clear communication (e.g. air traffic control) it is possible a PWS who is rejected at interview solely because of their stuttering may take legal action. There appears to be no studies in this area: namely the extent to which speech dysfluency on its own or together with other specific factors, influences occupational selection decisions.

5. Summary and conclusions

Social and emotional factors appear to be crucial to the onset and maintenance of stuttering. This is evidenced by the fact that most models explaining the onset and development of stuttering include a social and emotional component. This paper has reviewed the methods and findings of previous research that investigated the role of affective and social factors in stuttering from pre-school to adulthood. It has also attempted to make readers aware of the various methods and issues in social psychology that could be used to investigate these phenomena and to indicate where these methods could be useful in assessing the role of social and affective components in stuttering.

Acknowledgement.

This research was supported by the Wellcome Trust.

References

Andrews, G. & Cutler, J. (1974). Stuttering therapy: The relation between changes in symptom level and attitudes. *Journal of Speech and Hearing Disorders,* **39**, 312-319.

Andrews, G., and Harris, M. (1964) *The Syndrome of Stuttering: Clinics in Developmental Medicine 17*. London: Heinemann.

Anderson, J. D., Pellowski, M. W., Conture, E. G., & Kelly, E. M. (2003). Temperamental characteristics of young children who stutter. *Journal of Speech, Language, and Hearing Research*, **46**, 1221-1233.

Au-Yeung, J., Howell, P., Davis, S., Sackin, S., & Cunniffe, P. (2000). Introducing the Preschoolers Reception of Syntax Test (PROST*). Proceedings of conference on Cognitive Development, Besancon France.*

Bloodstein, O. (1987). *A handbook on stuttering.* Chicago: The National Easter Seal Society.

Boutsen, F. & Brutten, G. (1989). *Stutterers and nonstutterers: A normative investigation of of children's speech associated attitudes.* Unpublished manuscript, Southern Illinois University.

Brown, S.F. & Hull, H.C. (1942) A study of some social attitudes of a group of 59 stutterers. *Journal of Speech Disorders***, 7***,* 153-159.

Brutten, G.J. (1985). *Communication Attitude Test.* Unpublished manuscript. Carbondale: Southern Illinois University, Department of Communication Disorders and Sciences.

Brutten, G.J. & Dunham, S. (1989) The communication attitude test. A normative study of grade school children. *Journal of Fluency Disorders,* **14***,* 371-377.

Conture, E.G. (1996). Treatment efficacy: Stuttering. *Journal of Speech and Hearing Research*, **39**, S18-S26.

Costello, J,M. (1984) Treatment of the young chronic stutterer in, R. F. Curles & W, H. Perkins (Eds.) *Nature and Treatment of Stuttering*. San Diego: College-Hill Press.

Dalton, P. & Hardcastle, W.J. (1977) *Disorders of fluency and their affects on communication.* London: Edward Arnold.

Davis, S., Howell, P., & Cook, F. (2002). Sociodynamic relationships between children who stutter and their non-stuttering classmates. *Journal of Child Psychology and Psychiatry*, **43**, 939-947.

De Nil, L. (1999). Stuttering: A neurophysiological perspective. In N.B. Ratner & E.C. Healey (Eds.), *Stuttering research and practice: Bridging the gap.* Mahwah, NJ: Lawrence Erlbaum.

De Nil, L. & Brutten, G. (1986). Stutterers and nonstutterers: A preliminary investigation of children's speech-associated attitudes*. Tijdschrift voor Logopedie en Audiologie,* **16**, 85-92.

De Nil, L. & Brutten, G. (1991) Speech-associated attitudes of stuttering and nonstuttering children. *Journal of Speech and Hearing Research*, **34***,* 60-66.

Erikson, R. L. (1969) Assessing communication attitudes among stutterers. *Journal of Speech and Hearing Research,* **12**, 711-724.

Guitar, B. (1976) Pretreatment factors associated with the outcome of stuttering therapy. *Journal of Speech and Hearing Research,* **19***,* 590-600.

Johnson, W., Brown, S.F., Curtis, J.F., Edney, C.W., & Keaster, J. (1956*). Speech handicapped school children (revised edition)*. New York: Harper Row.

Niven, N. (1994) *Health Psychology*. London: Churchill Livingstone.

Okasha, A., Bishry, Z., Kamel, M., & Moustafa, M. (1974). Psychosocial study of stammering in Egyptian children. *British Journal of Psychiatry,* **124**, 531-533.

Onslow, M. (1994). The Lidcombe Programme for early stuttering intervention: the hazards of compromise. *Australian Communication Quarterly, Supplementary Issue,* 11-14.

Perry, D.G, Bussey, K. and Fischer, J. (1980). Effects of rewarding children for resisting temptation on attitude change in the forbidden toy paradigm. *Australian Journal of Psychology,* **32**, 225-234.

Peters, T, J. and Guitar, B. (1991) *Stuttering: An Integrated Approach to Its Nature and Treatment.* Maryland: Williams & Wilkins.

Riley, G., & Riley, J. (2000). A revised component model of diagnosing and treating children who stutter. *Contemporary Issues in Communication Science and Disorders,* **27**, 188-199.

Schindler, M.D. (1955). A study of educational adjustments of stuttering and non-stuttering children. In W. Johnson & R. R. Leutenegger (Eds). *Stuttering in children and adults.* Minneapolis: University of Minnesota Press.

Sheehan, J.G. (1970). *Stuttering: Research and therapy.* New York: Harper and Row.

Smith, A. (1999). Stuttering: A unified approach to a multifactorial, dynamic disorder. In N.B. Ratner & E.C. Healey (Eds.), *Stuttering research and practice: Bridging the gap.* Mahwah, NJ: Lawrence Erlbaum.

Starkweather, C. W. (1985). The development of fluency in normal children. In H. Gregory (Ed.), *Stuttering therapy: Prevention and intervention with children* (pp. 9-42). Memphis, TN: Speech Foundation of America.

Starkweather, C.W., Gottwald, S.R., & Halfond, M.M. (1990). *Stuttering prevention: A clinical method.* Englewood Cliffs, NJ: Prentice-Hall.

St Louis, K. O., & Lass, N. J. (1981). A survey of communicative disorders: Students' attitudes towards stuttering. *Journal of Fluency Disorders*, **6**, 49-79

Travis, L.E. (1957). *Handbook of speech pathology.* East Norwalk, CT: Appleton-Century-Crofts.

Van Riper, C. (1948). *Stuttering.* Chicago, IL: National Society for Crippled Children and Adults, Inc.

Vanryckeghem, M. (1995) The Communication Attitude Test: A concordancy investigation of stuttering and nonstuttering children and their parents. *Journal of Fluency Disorders*, **20**, 191-203.

Vanryckeghem, M. & Brutten, G, J. (1992) The Communication Attitude Test: A test-retest reliability investigation. *Journal of Fluency Disorders*, **17**, 109-118.

Vanryckeghem, M. & Brutten, G, J. (1996) The relationship between communication attitude and fluency failure of stuttering and nonstuttering children. *Journal of Fluency Disorders*, **21**, 109-118.

Vanryckeghem, M., Hylebos, C., Brutten, G.J., & Perelman, M. (2001) The relationship between communication attitude and emotion of CWS. *Journal of Fluency Disorders*, **26**, 1-15.

Wall, M.J., & Myers, F.L. (1984) *Clinical management of childhood stuttering. Austin, TX: Pro-Ed.*

Watson, J, B. (1987) Profiles of stutterers' and nonstutterers' affective, cognitive and behavioral communication attitudes. *Journal of Fluency Disorders,* **12**, 389-405.

Yairi, E., & Ambrose, N. (1992). Onset of stuttering in pre-school children: selected factors. *Journal of Speech and Hearing Research*, **35**, 782-788

Yairi, E., & Ambrose, N.G. (1999) Early childhood stuttering 1: Persistency and recovery rates. *Journal of Speech, Language and Hearing Research,* **42**, 1097-1112.

Yeakle, E., & Cooper, E. B. (1986). Teacher perceptions of stuttering. *Journal of Fluency Disorders*, **11**, 345-359.

Zimmerman, G. (1980). Stuttering: A disorder of movement. *Journal of Speech and Hearing Research,* **23**, 122-136.

RESEARCH COMMENTARIES ON FURNHAM AND DAVIS

The social phenomena of stuttering

Walter H. Manning, Ph.D
Memphis Speech & Hearing Center
807 Jefferson Avenue
Memphis, Tennessee 38105 USA
wmanning@memphis.edu
Website: http://www.ausp.memphis.edu/people/manning.html

Abstract. Social and affective factors appear to have greater influence on the development than the onset of stuttering. Although the dynamic characteristics of these many factors make them difficult to research they are important treatment issues. In most instances, the frequency of overt stuttering behavior and the speaker's affective and social response to stuttering, are different issues.
Keywords: Social factors in stuttering.

1. Comments on Furnham and Davis (2004)

I have little disagreement with Furnham and Davis (2004) concerning their interpretations and summaries of affective and social factors as important components of the stuttering experience. Because of the range and interactions of the factors discussed, the reader may struggle to find a focus in the article. Further, the distinctions between social and affective factors are not always clear. I also noted several sentences that were somewhat vague (e.g., "Change of attitude is especially likely to be the case when a child has a problem that affects overt behavior." Also on the same page, a reference for the "theory of planned action" would have been helpful.

Speaking more or less fluently involves the dynamic interaction of many affective and social factors which vary according to the participants. However, these factors play important roles in the development of stuttering. Persons who fully understand the phenomenon generally agree that these factors are better indicators of the trauma and handicap experienced by the speaker than measures such as the frequency of overt stuttering.

It's also important to point out that many of the affective and social aspects of stuttering don't necessarily coincide with an increase in fluency. Because it often takes some time and effort to achieve a paradigm shift about yourself and your possibilities, even the achievement of fluency may not lead to a more spontaneous and responsive lifestyle. Particularly for the adult who has stuttered and survived in the culture of stuttering for many years and it can be extremely difficult to let go of the affective and social responses that allowed you to survive.

From my understanding of the literature and from my personal and clinical experience, I don't believe that social factors influence the onset of stuttering, a popular view in the middle of the last century. Aside from the fact that these factors are sometimes included in models of onset, I don't believe that the evidence cited here (or in general for that matter) provides a convincing argument that social factors are crucial to stuttering *onset*. Social factors, as described by Furnham and Davis, do appear to be an important component in the development and maintenance of the problem. Most in the field of fluency disorders, even those with diverse views of what to do therapeutically, feel that the onset is governed by physiological predisposing factors (my favorite explanation being that of Smith and her colleagues (Smith & Kelly, 1997, Smith, 1999). But social factors certainly appear to help keep stuttering going and growing.

Although we know very little about the social response to stuttering in many cultures, every study of which I am aware has documented the idea that stuttering is more or less stigmatized by listeners of all cultures. This is one of the primary reasons people who stutter try so hard not to do so. This, of course, leads to problems and generally increases the overt struggle behaviors as well as the speaker's decisions about how to hide from the problem. As a result, PWS tend to make restricted decisions, greatly contributing to the handicapping nature of the problem. Paradoxically, to the extent that many PWS "give themselves permission" to stutter, they stutter with considerably less effort and often, not at all.

As Furnham and Davis indicate, intelligence does not appear to be a compelling factor for understanding stuttering. Basic personality characteristics of people who stutter are not profoundly unique. But, related to these personality factors is the interesting notion of temperament, an area likely

to lead to useful information for understanding the phenomena and for facilitating change during treatment, particularly with children. It may be that children come to stuttering with these characteristics. On the other hand, because we are now realizing that even very young children are aware of their predicament at much earlier ages than many had suspected (Yairi, 2004) it may be that the experience of stuttering influences temperamental factors even at a very early age.

Stuttering is both an obvious and easy target for bullies. Such predators seek easy targets and tend to run if challenged. Fortunately, there are many possibilities for adaptive responses and a teasing and bullying is a worthy target for the clinician's therapeutic strategies and techniques (Blood & Blood, 2004). As the authors point out, people who stutter can be limited vocationally. To the extent that it is possible, educating listeners including employers often makes the problem less mysterious and manageable. In spite of the fact that stuttering is more variable than any other human communication disorder, it tends to be easily stereotyped. More knowledgeable observers will find that it is difficult to predict how a PWS will respond to situations. Many PWS stutter during speaking situations associated with no apparent anxiety and achieve high levels of fluency in situations when it would not be predicted. Although life is more likely to be challenging on many levels when you are somebody who stutters there are many people who achieve high levels of accomplishment and acceptance even when some stuttering is present.

Although the findings and methodology of several older studies are mentioned, there are several recent articles by Yairi and his colleagues (2004) concerning the likelihood of recovery as well as teasing and bullying (see Blood and Blood, 2004) that are not included in this review (the 2004 article by Healey, et al. in this same volume was cited). Finally, in spite of the methodological problems described by Furnham and Davis, I believe that self-reports and even retrospective qualitative studies can contribute to our knowledge and understanding of the experience of stuttering. To the extent that we formulate the questions we are asking using seemingly more controlled and reliable measures in a series of questions or scales, we are more likely to introduce themes which influence participant responses.

References

Blood, G. W. & Blood, I. M. (2004). Bullying in adolescents who stutter: Communicative competence and self-esteem. *Contemporary Issues in Communication Science and Disorders*, **31**, 69-79.

Furnham, A. & Davis, S. (2004). Involvement of social factors in stuttering: A review and assessment of current methodology. *Stammering Research,* **1***,* 112-122.

Healey, E. C., Trautman, L. S., & Susca, M. (2004) Clinical applications of a multidimensional approach for the assessment and treatment of stuttering. *Contemporary Issues in Communication Science and Disorders*, **31**, 40-49.

Smith, A. (1999). Stuttering: A unified approach to a multifactorial, dynamic disorder. In N. B. Ratner and E. C. Healey (Eds.), *Stuttering Research and Practice: Bridging the Gap* (pp. 27-44). Mahwah, NJ: Lawrence Erlbaum.

Smith A., & Kelly, E. (1997). Stuttering: A dynamic, multifactoral model. In R. F. Curlee & G. M. Siegel (Eds.), *The Nature and Treatment of Stuttering: New directions* (2nd Ed. pp. 204-217). Needham Heights, MA: Allyn & Bacon.

Yari, E. (2004). The formative years of stuttering: A changing portrait. *Contemporary Issues in Communication Science and Disorders*, **31**, 92-104.

Extending the scope of stuttering research and treatment

John Wade
Counseling and Psychological Services, University of Kansas, 1200 Schwegler Drive, Lawrence, KS 66045, USA
jwade@ku.edu

Abstract. The target article reviews the assessment and research methodology of social and emotional factors involved in stuttering. I offer comments based from my perspective as both a psychologist and also as a person who stutters, about the need to view both research and treatment of stuttering from a holistic perspective. This would include expanding both research and treatment to include a greater focus on communication effectiveness, "quality of life issues," and researching individuals who deal successfully with their stuttering.
Keywords: Listener reaction, self-perception, coping skills, quality of life.

1. Comments on Furnham & Davis (2004)

Furnham and Davis' (2004) target article addresses the important topic of conducting relevant and meaningful research on social and emotional factors that impact stuttering across the life-span. As with most fields, it is likely that a holistic approach to both research and treatment, involving other disciplines including psychology, sociology, and communication studies is likely to be the future trend, and is likely to yield the most benefit.

Since stuttering is a communication disorder, both listener reactions and self-perceptions play a vital role in determining the functional impact of the individual's stuttering on his or her communication and overall quality of life. Klassen's (2001) review of past research indicates that people who stutter (PWS) are perceived in a number of socially undesirable ways. Although both clinical impressions and research findings regarding PWS' self-perception are somewhat mixed (e.g., Shames & Rubin, 1986; Blood, Blood, Tellis & Gabel, 2003), and much of the research is dated, it seems clear that the impact of listener reactions can have a profound impact on PWS self-perceptions. However, usefulness is usually found in specifics. It is not enough to simply know that stuttering may or may not be negatively perceived, or even to know the different negative attributes that may be associated with stuttering. More research needs to be conducted that focuses on determining exactly what specific aspects of stuttering are perceived negatively by listeners. Are there ways to stutter that engender more positive listener reaction and greater comfort with disfluent speech? As Wendell Johnson stated, we may not be able to determine whether we stutter but we can change how we stutter.

Achieving greater fluency is obviously an important component of communicating effectively, but it is only one of several factors that contribute to successful communication. Kamhi (2003) and others have asserted the need for speech therapy to focus on communication, not just stuttering. There is a wealth of literature in communication studies (e.g., Gabor, 1997; Garner, 1997; Patterson et. al 2002) that provides information on communicating effectively and making positive interpersonal impressions. This information needs to be incorporated into both research, and most importantly, treatment of stuttering.

A common criticism of psychology, which is my discipline, is that too great an emphasis has been placed on studying dysfunction. Recently, the field of positive psychology has emerged which focuses on learning about optimal psychological health and functioning by studying individuals at the highly successful end of the continuum. This same approach needs to be applied to research on stuttering. PWS who elicit positive listener reactions and those who have positive self-esteem and low emotional distress related to their stuttering need to be identified and studied. Important questions need to be addressed, such as "what behaviors do PWS who are perceived positively by listeners display?," "what are the pathways people who deal successfully with their stuttering utilize?," and "what coping skills do PWS who experience only low levels of communication frustration employ that PWS who experience high levels of frustration do not?" Examining social and emotional variables is important for understanding the onset and development of stuttering, but it is probably even more important for gaining insight into effective treatment and for developing effective coping skills.

Two years ago, the National Stuttering Association sponsored a research symposium, where more than 60 researchers, clinicians, and consumers met to discuss areas for future research in the field of stuttering. I participated in the sub-group that discussed future directions of research regarding stuttering treatment. The majority of the discussion centered on the consensus that meaningful assessment of therapeutic success must include assessment of communication effectiveness, emotional and interpersonal functioning, and general "quality of life" issues. Numerous measures to assess changes in these areas have been developed and could readily be utilized or modified to more specifically apply to stuttering and the related social and emotional factors.

The discussion in the Furnham and Davis (2004) article, of the negative perceptions listeners tend to have of stuttering, illuminates the importance of education. Depending on the level of severity, stuttering often is not a problem simply because the speech flow is disrupted or because it takes longer to say something. It is problematic, at least in part, because of the negative stereotypes that exist regarding stuttering and consequently the negative attributions made regarding PWS. One of the powerful experiences people who attend the National Stuttering Association annual convention usually experience is the comfortableness of spending three or four days interacting with people who understand stuttering and do not hold negative stereotypes based on level of fluency. It is important to educate the general public and provide accurate information to replace the negative inferences that can arise from the lack of knowledge. It is also important to provide accurate information about stuttering to PWS, who are not immune from having negative stereotypes of stuttering and consequently negative self-perceptions regarding their speech.

References

Blood, G.W., Blood, I.M., Tellis, G.M., & Gabel, R.M. (2003). A preliminary study of self-esteem stigma, and disclosure in adolescents who stutter. *Journal of Fluency Disorders,* **28**, 143-159.

Furnham, A. & Davis, S. (2004). Involvement of social factors in stuttering: A review and assessment of current methodology. *Stammering Research,* **1***,* 112-122.

Gabor, D. (1997). Talking with confidence for the painfully shy. New York: Random House.

Garner, A. (1997). *Conversationally speaking: Tested new ways to increase your personal and social effectiveness*. Los Angeles: Lowell House.

Kamhi, A.G. (2003). Two paradoxes in stuttering treatment. *Journal of Fluency Disorders,* **28**, 187-195.

Klassen, T.R. (2001). Perceptions of people who stutter: Re-assessing the negative stereotype. *Perceptual and Motor Skills,* **92**, 551-559.

Patterson, K. Grenny, J., McMillan, R., & Switzer, A. (2002). *Crucial conversations: Tools for talking when the stakes are high.* New York: McGraw Hill.

Shames, H., & Rubin, R. (1986). *Stuttering then and now.* Columbus, OH. Merrill Publishing Co.

Commentary on "Involvement of social factors in stuttering: A review and assessment of current methodology" by A. Furnham and S. Davis

O. P. Skljarov
Research Institute of ETN & Speech, 190013, St.Petersburg, Bronnitskaja, 9, Russia
skljarov@admiral.ru

Abstract. Furnham and Davis' (2004) article reviews the methods and findings of previous research that investigated the role of affective and social factors in stuttering from pre-school to adulthood. Attention is drawn to Vinarskaja and Bogomazov's age-specific phonetics approach that provides an integrated framework for examining language and social factors.
Keywords: Affective and social factors, stuttering, paralinguistic approach, age-specific phonetics.

1. Introduction

The target article reviews the methods and findings of previous research that investigated the role of affective and social factors in stuttering from pre-school to adulthood. There are two main intentions in the target article.

The first is to present a review of methods and findings of previous research that investigated the role of affective and social factors in stuttering. The authors made a fine review, and presented many results obtained from a variety of different psychological tests. However, Russian papers in this area were not included. One Russian publication relevant to Furnham and Davis' topic area is the excellent book of Vinarskaja and Bogomazov entitled "Age-specific phonetics" and published in 2001. The theoretical premise of this book is that when assessing language it is essential to examine in a hierarchically-constructed system paralinguistic, emotionally-indicative and phonetic aspects. Many statements in this book are consistent with the findings of Skljarov (2004, submitted) and the findings of studies presented in the target article.

The second intention in the target article is to alert readers to various methods and issues being applied in social psychology to investigate stuttering and to indicate where these methods could be used when assessing the role of social and affective components in stuttering. Review of the various methods and issues is very complete and it will undoubtedly be of major interest for investigators of affective and social factors in stuttering. The methods used in age-specific phonetics is an omission. The method of approach is that the child does not acquire the adult language of a society. In each period of development a child adopts age-specific 'languages' that are formed as a result of normative social influences at each stage.

References

Furnham, A. & Davis, S. (2004). Involvement of social factors in stuttering: A review and assessment of current methodology. *Stammering Research,* **1***,* 112-122.

Skljarov, O.P. (2004). Speech ontogenesis and scenario of its V-rhythms. *Electron Journal Technical Acoustics" http:/webcenter.ru/~eeaa/ejta,* **7**.

Skljarov, O.P. (submitted). Duality of the Feigenbaum's scenario for V-rhythm of speech. Extending the Scope of Stuttering Research and Treatment *Stammering Research*.

Vinarskaya E.N., & Bogomazov G.M. (2001). *Age-specific phonetics*. STT: Tomsk.

AUTHORS' RESPONSE TO COMMENTARIES

Authors response to commentaries on 'Involvement of social factors in stuttering: A review and assessment of current methodology'

Stephen Davis and Adrian Furnham
Department of Psychology, University College London, Gower St., London WC1E 6BT, England
stephen.davis@ucl.ac.uk
a.furnham@ucl.ac.uk

1. Introduction

The authors are grateful to the commentators who responded to our target article (Furnham & Davis, 2004). We were pleased that the commentaries were appreciative of the review and that there was a consensus between the commentators regarding the importance of social factors in stuttering research. The authors welcome the opportunity to address a few points made by the commentators.

2. Positive psychology

The authors welcome Dr Wade's (2004) introduction of positive psychology into the review of social factors. Although the review of past research that we undertook generally revealed negative concepts of people who stutter there is a role for positive psychology in both research and clinical practice. Positive psychology is one approach for investigating the relationship between predictor and outcome variables. Along with Dr Petrides and Professor Howell, we are examining a range of different statistical approaches (over and above positive psychology) to these same phenomena.

3. Social factors and onset

Although we generally agree with Dr Manning's (2004) position that social factors play little part in the onset of stuttering there are indications that young children are aware of their stutter close to onset (Davis et al, in press; Yairi 2004). It would not be unreasonable to conclude that these young children would also soon be aware of the negative reactions that their stutter could induce from their peers. This has the potential to make the child who stutters more reticent to engage in social contact. In turn this could lead to the rejection, isolation and lowered self-esteem that are implicated in teasing and bullying (Hodges & Parry, 1996). The possibility that young children are aware of their stutter close to onset underlines the importance of social factors in monitoring and predicting the developmental pathway of childhood stuttering.

4. Personality and temperament

Dr Manning also proposes that temperament may provide useful information allowing clinicians to facilitate change during treatment. Eastern European theories of temperament (eg The Regulative Theory of Temperament (Strelau, 1996)) suggests that temperament is biological and that temperament characteristics are present from early infancy. This would indicate that temperamental characteristics are not only useful in a therapeutic environment but should also be included in models that include a physiological component in predicting the onset of stuttering (Smith, 1999; Smith & Kelly 1997).

5. Communication attitude

Although Dr Vanryckegham was unable to submit a full commentary she has pointed out the importance of monitoring the communication attitudes of young children who stammer. The authors fully agree with this position and are pleased to note the work currently in progress in this area developing instruments designed to measure the communication attitudes of children aged 3-7 years (Davis et al, in press; Vanryckeghem & Brutten, 2002; Vanryckeghem, Hernandez & Brutten, 2001).

References

Furnham, A. & Davis, S. (2004). Involvement of social factors in stuttering: A review and assessment of current methodology. *Stammering Research,* **1***,* 112-122.

Davis, S., Howell, P., Killick, A., & Finch, H. (in press). The development and validation of a communication attitude instrument for use with young children. *Proceedings of the 4th World Congress on Fluency Disorders,* Montreal: IFA.

Hodges, E. V. E., & Parry, D. G. (1996). Victims of peer abuse: An overview. *Journal of Emotional and Behavioural Problems,* **5,** *23-28*

Manning, W. H. (2004). The social phenomena of stuttering. *Stammering Research,* **1***,* 123-124.

Skljarov, O. P. Commentary on "Involvement of social factors in stuttering: A review and assessment of current methodology" by A. Furnham and S. Davis. *Stammering Research,* **1***,* 127.

Smith, A. (1999). Stuttering: A unified approach to a multifactorial, dynamic disorder. In N. B. Ratner and E. C. Healey (Eds.), *Stuttering Research and Practice: Bridging the Gap* (pp. 27-44). Mahwah, NJ: Lawrence Erlbaum.

Smith A., & Kelly, E. (1997). Stuttering: A dynamic, multifactoral model. In R. F. Curlee & G. M. Siegel (Eds.), *The Nature and Treatment of Stuttering: New directions* (2nd Ed. pp. 204-217). Needham Heights, MA: Allyn & Bacon.

Srelau, J., (1996). The regulative theory of temperament: Current status. *Personality and Individual Differences,* **20**, 131-142

Vanryckeghem, M & Brutten, G (2002). KiddyCAT: A measure of stuttering and nonstuttering preschooler's attitude. *ASHA Leader*, **7**, 104.

Vanryckeghem, Hernandez & Brutten (2001). The KiddyCAT: A measure of speech associated attitudes of preschoolers. *ASHA Leader*, **6**, 136.

Wade, J. (2004). Extending the scope of stutteirng research and treatment. . *Stammering Research,* **1***,* 125-126.

Yairi, E. (2004). The formative years of stuttering: A changing portrait. *Contemporary Issues in Communication Science and Disorders*, **31**, 92-104.

RESEARCH DATA, SOFTWARE AND ANALYSIS SECTION

Facilities to assist people to research into stammered speech

Peter Howell1[1] and Mark Huckvale[2]
[1]*Department of Psychology, University College London, Gower St., London WC1E 6BT*
p.howell@ucl.ac.uk

[2]*Department of Phonetics and Linguistics, University College London, Gower St., London WC1E 6BT*
m.huckvale@phon.ucl.ac.uk

Abstract. The purpose of this article is to indicate how access can be obtained, through *Stammering Research*, to audio recordings and transcriptions of spontaneous speech data from speakers who stammer. Selections of the first author's data are available in several formats. We describe where to obtain free software for manipulation and analysis of the data in their respective formats. Papers reporting analyses of these data are invited as submissions to this section of *Stammering Research*. It is intended that subsequent analyses that employ these data will be published in *Stammering Research* on an on-going basis. Plans are outlined to provide similar data from young speakers (ones developing fluently and ones who stammer), follow-up data from speakers who stammer, data from speakers who stammer who do not speak English and from speakers who have other speech disorders, for comparison, all through the pages of *Stammering Research*. The invitation is extended to those promulgating evidence-based practice approaches (see the *Journal of Fluency Disorders, volume 28, number 4* which is a special issue devoted to this topic) and anyone with other interesting data related to stammering to prepare them in a form that can be made accessible to others via *Stammering Research*.
Keywords: UCL Psychology speech group database http://www.speech.psychol.ucl.ac.uk/, SFS http://www.phon.ucl.ac.uk/resource/sfs/, CHILDES http://childes.psy.cmu.edu, PRAAT *http://www.praat.org*, the Wellcome Trust *http:wellcome.ac.uk*.

1. Introduction

Over the last 20 years, spontaneous speech data from speakers who stammer have been collected by the Speech Group in the Psychology Department of University College London (UCL). These data have mainly been collected through research funding from the Wellcome trust who, as a matter of policy, encourage public access to science. A proportion of these data has been transcribed. All these materials will be shared with the research community with the intention of encouraging research into stammering. In the future, we plan to make other types of data available provided by our own, and hopefully other, groups. Some background considerations about the choice of approach to make these data available are given in section 2 of this article.

Section 3 includes a description of the data we hold, and indicates which subset we are currently making available. The transcribed data are available in CHAT, PRAAT TextGrid and (in some cases) as SFS annotation items aligned against the audio records. The audio data have been prepared in WAV, SFS and MP3 formats. Undoubtedly there are other packages available too and authors or users of those systems are welcome to prepare the data so that they can be processed with them. Users can submit articles that make the case for including other formats. If such articles are accepted after peer review, they will be invited to prepare the data in that format and it will be included in the archive.

Providing samples of data from speakers who stammer is an important step in encouraging research into stammering. However, people wishing to do research also need to have facilities to manipulate these data (for analyses of speech characteristics, to use these data in perceptual assessments and so on). The formats that are provided allow readers who are familiar with CLAN, PRAAT and SFS programs to investigate these data. Section 4 gives tutorial material and some illustrative analyses using the SFS software suite. SFS is used by the UCL speech group for the reaons also given in section 4. It is necessary to emphasize that exclusive use of SFS is *not* being advocated; CHILDES and PRAAT each have facilities that are not available in SFS (and vice versa). Thus, different software packages should be chosen according to what analyses end-users wish to perform. As indicated in the previous paragraph, some preparation has been done on the data so that each of these software packages can process the data. Future issues of *Stammering Research* will include details and demonstration of some of the capabilities of these other software suites. A general feature that commends SFS, CHILDES and PRAAT, is that they are all available for free. Details of how to access these software are included.

In addition to data and software for analyzing these data, researchers need an outlet for their findings. It is suggested that researchers consider *Stammering Research* as such an outlet. This will allow an archive of analyses to be built up on data available through the journal, provide a forum to

make corrections to the data, and offer a repository that can be used to access new software that are useful for analysis of data like these.

This is an extensive document. It has been written as far as is possible so that the different sections can be read in isolation. The body of the text gives the description of what data and software are available and also reports studies using these data that we hope have some intrinsic interest to readers. The material in the Appendices will need to be consulted when a reader has some specific need. Appendix A will be used when the user wishes to select some of the data in whatever format they need. Appendix B describes the machine-readable transcription convention that UCL Psychology's Speech Group has adopted which users may choose to use or convert to one they are more familiar with. Appendices C and D give worked exercises using SFS utilities in two important areas for manipulation of these data (transcription and formant analysis). Appendix E gives some details of applications of a Hidden Markov model software suite to recognition of dysfluencies. The background knowledge that is assumed is given at the start of each Appendix. These appendices were, in part, written as tutorial material for this article though they also serve as important sources for general users of SFS. Also, applications in the body of this paper draw on the information provided in these appendices. For all these reasons, it is appropriate that they be included in this document. It should be apparent that these facilities are only part of what SFS offers. There is extensive other software in and various options for people wishing to write their own scripts (in MATLAB, C, and in speech measurement language, SML which is a high level scripting language for manipulating information in SFS files, for details of SML see section 1.5 of the SFS manual at http://www.phon.ucl.ac.uk/resource/sfs/). The CHILDES and PRAAT systems have various similar options that, it is hoped, will be described in future articles.

2. Background to release of data on speakers who stammer and software for its analysis

Speech data are expensive to collect, speech is time-consuming to analyze, the range of analyses that one group can do is limited (e.g. because of time constraints). Individuals who want to start a research program into stammering often have the daunting tasks of obtaining expertise, equipment, software and an administrative structure (to locate participants, obtain ethical permission etc.) before they can conduct their research. If they are interested (like us) in developmental issues associated with stammering, they may have to collect longitudinal data for several years.

There are pitfalls when attempting to make recordings. Clinics and schools are not ideal recording environments. (Many recordings that speech therapists have offered us in the past have been too poor in quality for analysis.) Special skills need to be acquired for eliciting speech from young speakers and also some children who stammer.

UCL Psychology Department's speech group has extensive audio data on speakers who stammer of appropriate quality for linguistic and acoustic analysis, which it is prepared to supply to the wider community thereby circumventing these problems in 'getting started' in research into stammering. Generally speaking, three issues need to be addressed so the data can be used. These are 1) conventions for data preparation need to be specified, 2) information needs to be supplied about how to use available software to manipulate data in these formats, and 3) information has to be provided about where and how other similarly formatted data can be deposited or accessed.

The way in which three different systems address these issues is discussed. The three systems selected for consideration are MacWhinney's CHILDES system, Boersma's PRAAT system and Huckvale's SFS system. Each of these systems was developed for different purposes and each has advantages that make it more useful for some purposes than are the others. This is implicitly acknowledged as, for instance, there is provision in the CHILDES system to access PRAAT software to take advantage of the facilities for speech analysis provided by PRAAT. The components of each system and the purpose for which they were developed are briefly indicated.

CHILDES. The CHILDES project was developed specifically for research into child language. It is masterminded by Brian MacWhinney and has three separate components that address the above three issues: CHAT (Codes for the Human Analysis of Transcripts) provides the data conventions, CLAN (Computerized Language ANalysis) is the software package and CHILDES (CHIld Language Data Exchange System) is the repository for the data. The three components together are referred to as the CHILDES project (MacWhinney, 1995). CHILDES was developed for higher level linguistic analysis. This makes sense in terms of the target populations for which the system was developed. Lengthy recording sessions are often necessary to obtain even limited samples of speech from very young children. Consequently, it would not make sense to archive the entire audio recordings in these cases (a lot of the data so stored would be taken up by interlocutors and sections where the child says nothing). The system has been developed so that audio data can be logged and there are links to audio analysis software (e.g. PRAAT). This is particularly useful when samples from older speakers are employed. Most of the data that are available through CHILDES are from fluent speakers. Further details can be found on the CHILDES home page at http://childes.psy.cmu.edu.

PRAAT. PRAAT was developed by Paul Boersma and is specifically an audio data analysis tool and PRAAT has the advantage that it is simple to use. It reads all kinds of standard audio format files including WAV. It runs on Unix, Mac and Windows platforms. Transcription data are in the form of a TextGrid which can easily be created within the PRAAT system or imported via a suitably formatted text file. The software incorporates sub-tools for neural net analysis, speech synthesis and some statistical functions, and a scripting facility is available which allows users to write specialized functions. There is a PRAAT user group who share data (as well as software). PRAAT is close to SFS in the facilities it provides, but each of them have specialized processing software. For example, PRAAT does sophisticated analyses of harmonicity while SFS incorporates software for dealing with ancillary signals provided from a laryngograph . The PRAAT Home Page is at http://www.praat.org and the manual is accessed within the software.

SFS. SFS stands for Speech Filing System. It provides an integrated method of dealing with different sources of information about speech sounds. The raw audio record is at the core of the system. There are options so that a number of other analyses (manual or computational) on the same speech data can be displayed for inspection. Transcriptions can be manually entered in any format (Appendix B gives the Joint Speech Research Unit format, JSRU, that the UCL Speech Group uses). The filing system provides the system that integrates analyses from these several sources for visual or statistical inspection. The integration of these sources of information is the attraction, though there are utilities that allow the audio recordings to be uploaded or dumped in WAV or other standard formats and, similarly, TXT files can be dumped or uploaded.

For the Speech Group's work, the SFS facility that allows, *inter alia*, audio data and aligned transcriptions to be concurrently displayed has been particularly useful in the development of Howell's EXPLAN theory of spontaneous speech control (Howell, 2002, 2004; Howell & Au-Yeung, 2002). This theory maintains that motor execution of one segment takes place concurrent with the planning of the following segment and that fluency may break down when execution time on one segment does not allow sufficient time to complete planning of the following segment. SFS displays of the audio waveform and the associated transcription provide the information necessary for evaluation of predictions of this theory. Thus, the audio item provides an indication of the time required for execution of the current segment, the annotation item indicates the structure of the following segment that can be used to ascertain how complex the word is to plan (Dworzynski & Howell, 2004; Howell & Au-Yeung, 1995a; Howell, Au-Yeung & Sackin, 2000). These two factors can be examined jointly to determine whether they lead to fluency breakdown.

The software provided by SFS includes many of the same facilities as PRAAT. SFS can display selected analyses of the same stretch of speech aligned in time. Appendix A of this article gives details of how to access an extensive SFS database (i.e. the data on speakers who stammer). SFS was developed on a Unix system, and certain advanced software features require use of the Unix command line interpreter (such as those in Appendix E). The SFS Home Page can be found at http://www.phon.ucl.ac.uk/resource/sfs/

Although our work employs SFS extensively, other users may prefer to work with one of these other systems if they have specific needs of the facilities provided. Basic versions of the files have been supplied that allow users of these other systems to get started. For CHILDES users, files have been converted according to TEXTIN (MacWhinney, 1995, p.158), and links to WAV files provided. Other information that could be added to the CHAT files (speaker, age etc.) are available in the Access file described in Appendix A. The files have also been prepared as PRAAT TextGrids (and WAV files) at Paul Boersma's request. He intends to inform his email list of PRAAT users about the possibility of analysing and reporting results on these data. It is also recognised that some users may just need to listen to the files or read the transcriptions. The transcriptions can be examined with current word processing packages. Users who do not intend to carry out acoustic analyses probably do not want hi-fidelity WAV files as these are cumbersome to access. For this reason, MP3 files have been made available which are much shorter files. Though there is some data loss, the audio files are still of good quality. Most of the remainder of this article refers to the SFS versions of the files and associated analysis software.

Access to UCL Psychology Department Speech Group's data files. This article is intended to set the ball rolling to stimulate research in this area. Part of UCL's data on speakers who stammer will be made available. We are not just going to supply the audio data, but also, where available, orthographic and phonetic transcriptions. Some of the latter are also aligned against the audio records.

Appendix A indicates how the data in the various formats can be accessed. The data will also be distributed in an alternative medium that some users may find more convenient. CDs have been made of different subsets of the data set (described in Appendix A). The listening center at UCL's Phonetics and Linguistics Department holds copies of these CDs (the data cannot be customized for individual clients so if you need selections of data appearing on two or more CDs, you will have to purchase the

CDs concerned). CDs can be purchased for a cost of around £10 each (including p&p) for those who prefer their data in this format. To comply with the European Union's data protection act, we have ensured that speakers cannot be identified from the recordings.

Transcription of spontaneous speech is a difficult task and it is unlikely to lead to 100% agreement between transcribers, and the transcription procedure has been designed to be more detailed at some points in utterances (around dysfluencies) than others (the rest of the speech). The reliability estimates that have been made for a selection of the transcriptions indicate that there is satisfactory agreement between trained transcribers (Howell, Au-Yeung & Sackin, 1999, 2000). However, this is the first time these transcriptions have been open to public scrutiny and although we have been rigorous in preparing the transcriptions, there will inevitably be some errors. Users should notify by email to psychol-stammer@ucl.ac.uk any errors they locate so that these can be corrected in subsequent release versions of the data. This 'public' correction procedure is preferred to one where audio files are not available, so errors are not visible. In addition, the recordings where there are audio files alone can be used by anyone who wishes to start from scratch (using the current or any other transcription scheme). We consider it imperative that new and improved transcriptions should be made available for scrutiny in the same manner as those we have provided.

Access to UCL's Phonetics Department's SFS software. SFS can be obtained from http://www.phon.ucl.ac.uk/ under Research Resources.

We do not have the resources to offer software support. We believe that providing these facilities offers a forum for scientific collaboration and exchange of ideas, which is the ethos behind *Stammering Research*. Copyright to the data is held by Howell and copyright to the software is held by Huckvale. The data and software are freely available to anyone for research and teaching purposes. If the data and/or software are used in publications, theses etc., users have to a) notify Howell (p.howell@ucl.ac.uk), b) acknowledge the source in any publication by referencing this article, c) include an acknowledgement that data collection was supported by the Wellcome Trust.

Outlet. It is intended that publications reporting analyses of these data will appear in *Stammering Research*. The prospect of extending research and assisting beginners to get started in research was made possible with the advent of Internet publications. In the main, the Internet has failed to deliver these possibilities to date because, where e-journals have appeared, they have usually been electronic versions of the printed journals that were available previously and have not provided access to data sources. *Stammering Research* welcomes submissions of articles for consideration that report analyses of these data, comparison between these data and previously published findings and so on. Submissions are invited at any time. There are no restrictions about what these analyses can be nor who may submit their work: Acoustic, articulatory, phonetic, phonological, prosodic and syntactic analyses would all be appropriate. The reports could also cover type/token, qualitative, transactional analyses. As implied, they can be compared with fluent speech or with samples of speech from people with other disorders or just analyses reporting on characteristics of stammered speech. Authors should be prepared to return analyses and scripts back to the archive (email submissions to psychol-stammer@ucl.ac.uk).

The SFS system can also be used for preparing material for perceptual tests. For instance, the software and the data could be used to replicate the classic Kully and Boberg (1988) study that showed that interclinic agreement in the identification of fluent and stammered syllables was poor. They can also be used to check some of the claims made by Cordes-Bothe and Ingham in support of time interval analysis as a means of assessing stammered speech. In this connection, it would be particularly useful to have TI sections of these freely-available materials judged by some members of these authors' expert panel, as the judgments of this panel have previously been used as benchmarks for other data about which intervals are, and are not, stammered.

It is hoped that these data will be of some lasting value to the research community (in the areas of stammering research and speech in general). In order to gauge whether there is a call for a facility like this, we have prepared a limited selection of our data set at present (more will follow if this proves popular). As stated above, a section of *Stammering Research* has been set up which is devoted to analyses that include (though is not necessarily restricted to) these data.

3. Description of the data

A complete description of the UCL archive of data from speakers who stammer is given and then, the subset in the initial release is described. The complete data set currently includes 249 speakers who are categorized into five classes depending on the range of ages over which they have been recorded. There is also a holding class for young children we are still seeing but cannot project how long they will be available for recording. The data within each of these classes have undergone different amounts of preparation levels. This document describes the version one release of data from the first class where recordings are only available over a limited age range and where there is no possibility of obtaining more recordings. Samples are a) available as audio alone (as SFS files), b) with orthographic

transcriptions (separate TXT file), c) with phonetic transcriptions (again separate TXT files) and d) where phonetic transcriptions are aligned against audio waveforms (available as SFS files). (CHILDES and PRAAT versions of the files are also available.) The alignment step under d) has a final check at the point where the transcriptions are aligned against the audio waveforms so these represent our highest level of data preparation. A full description of all classes of data we hold and current level of preparation is given below. The procedure for revising data in later revisions (corrections and generally useful ancillary analyses) is to send these in (as indicated in section 2). Depending on demand, other data classes will be released in phases as work is completed.

We record all speakers who stammer who volunteer. For our current project work, we are particularly interested in speakers in the age range eight to teenage. Pre teen speakers who stammer have a good chance of recovery. Consequently, we wish to follow up children who stammer and controls over this period, examine their speech and see whether different paths of fluency development are followed by those children who persist and those who recover. Ideally we want a minimum of three samples in this period (one in the age range 8-10 years, one between 10 and 12 years and one at teenage). We have complete sets of such recordings for 24 of our speakers.

Class 1, includes speakers who were either a) older than the maximum age of our target group when they were first seen (i.e. only seen after they have reached teenage), b) speakers who are in the target age ranges but who were only available at one target age because they live too far away from the laboratory or c) speakers who were in the required age range but with whom we have lost contact (most often because they have moved home and have not notified us of their new address and telephone number). Permission for data release cannot be obtained for speakers in class c). c) represents attrition of the sample, though there is no reason to suppose that this affects those speakers who persist differentially relative to those who recover. Class 2 consists of speakers who have not reached teenage who have been recorded at all target ages they have passed through that we are continuing to see (this class includes children who are under 8 years). Class 3 and 4 have not provided data at one of the first two target ages but attended at the third target age (i.e. they have reached teenage). For class 3, recordings are available at 8 and teenage, but not age 10-12 (often because the recording sessions clashed with school or family obligations and could not be rescheduled). For class 4, we have recordings for 10-12 and teenage. The lack of recordings at age 8-10 reflects the fact that these children were not seen at clinic until they were aged 10+. Class 5 are participants for whom we have at least three recordings at the designated ages, and we have continued to see most of these beyond the stipulated upper age. Many have supplied other forms of data. Appendix A describes an ACCESS file that is included in the data directory, which gives demographic information about the speakers.

Table 1 gives an indication of what is available and in what form. Not all data can be released at present (we have ethics permission, but are still waiting written consent by individuals or by clinics). Also, there are some data where the audio quality is not good enough for release. The numerator in each cell indicates what is being released and the denominator the total available. Thus 41/158 in the participants column indicates that data from 41 participants out of a total of 158 are being released.

Table 1. Summary of recordings available on UCL Department of Psychology Speech Group's data archive. Recordings are indicated according to class (see text) and type of file available.

	No of participants	No of files	No where orth avail	No where phon avail	No where phon aligned against audio
Class 1 –release 1	41/158	95/426	19/42	17/34	11/11
Class 2 - speakers who have not reached teenage, who have been recorded at all target ages they have passed through that we are continuing to see plus children who are < 8 years we are following up	7/ 21	13/57	5/7	1/1	2/2
Class 3 – available at 8-10yrs and 12yrs+	1/5	3/21	0/14	0/13	0/0
Class 4 – available at 10-12yrs and 12yrs+	7/41	20/184	6/57	6/54	0/1
Class 5 – recordings at the 3 target ages	5/24	7/142	1/63	0/44	3/69

	249	830	183	147	82
Totals	61	138	31	24	16

4. Some uses for the data including illustrations of applications of SFS tools and concepts for assessing stammered speech

Data similar to those made available in Appendix A have been used to investigate a range of questions about stammering from many different perspectives. A comprehensive list of all studies conducted is beyond the scope of this article. Studies conducted by the UCL group range from acoustic analysis of articulatory features associated with stammering, to pragmatic analyses of speakers who stammer in conversation with others. At the acoustic level, the UCL group has examined how the phonation source operates in people who stammer (Howell, 1995; Howell & Williams, 1988, 1992; Howell & Young, 1990), whether the vowel in a series of repetitions is neutralized using formant frequency analysis methods similar to those described in Appendix D (Howell & Vause, 1986) and speech rate has been measured from digitized oscillograms (Howell, Au-Yeung & Pilgrim, 1999; Howell & Sackin, 2000). PRAAT offers acoustic analysis software that could extend our understanding of what happens to the voice when fluency breaks down. Phonetic and phonological analyses have been performed to assess whether these factors are implicated in stammering (Dworzynski & Howell, 2004; Dworzynski, Howell & Natke, 2003; Howell & Au-Yeung, 1995a; Howell, Au-Yeung & Sackin, 2000). The change in pattern of stammering over development has been examined within prosodically-defined units in a variety of languages (Au-Yeung, Vallejo Gomez & Howell, 2003; Dworzynski, Howell, Au-Yeung & Rommel, 2004; Howell, Au-Yeung & Sackin, 1999) and different ways of defining these units (based on lexical or metrical properties) have been investigated for Spanish (Howell, in press). Various forms of syntactic analysis have been performed to establish whether syntactically complex utterances are more likely to be stammered than simpler ones (Howell & Au-Yeung, 1995b; Kadi-Hanifi & Howell, 1992). A pragmatic factor that has been examined is whether the speech of the interlocutor affects the speech of the person who stammers (Howell, Kapoor & Rustin, 1997). CHILDES offers a variety of techniques that extend the possibilities of examining other high order effects on stammering (including pragmatic ones). There is considerable scope for further phonetic, phonological, prosodic, syntactic and pragmatic analysis of these data and some suggestions follow.

Suggested studies

Perception of stutterings. The data that are supplied can be used for assessing the effect of different perceptual procedures on stammering assessment, for training therapists/pathologists on stammering assessments, for showing how heterogeneous stammering patterns can be within and across age groups. The materials could also be used to replicate the classic, but somewhat dated, Kully and Boberg (1988) study that showed judgements about the same sample of speech is judged differently by different clinics.

Studies on speech control in speakers who stutter. Some basic familiarity with acoustic phonetics is assumed to understand this section (for those requiring a refresher, see Ladefoged, 1975). The stammered sequence "kuh, kuh, Katy" contains a different sounding vowel ("uh" or as it is known more precisely "schwa"). Van Riper argued that a speaker who produces such a sequence had selected the wrong vowel at the start of this sequence and detected this by listening the sound of his or her voice (called feedback monitoring). As the speaker cannot produce "Katy" when the incorrect ("schwa") vowel has been inserted, the speaker interrupts speech and tries again. Howell and Vause (1986) argued that the vowel in a sequence of repetitions might sound like schwa because the vowels are short and low in amplitude (by analogy with vowel reduction that occurs in rapidly spoken, or casual, speech where the vowels also sound like schwa even when some other vowel is intended). They tested this hypothesis by acoustic analyses that compared the vowels in a sequence of repetitions with the vowel after fluent release (Howell and Vause also conducted perceptual tests, see the preceding section, which are not discussed here). They found that the formants of the vowels in sequences of repetitions and after fluent release occurred at the appropriate frequencies (suggesting that the vowel in the sequence of repetitions had been correctly articulated). Thus, they concluded that van Riper's feedback monitoring account of part word repetitions was not correct. The recordings of speaker 210 (at age 11 years 3 months) can be used to check Howell and Vause's finding. At 20.4s, the speaker appears to say "guh-go". The values of F1, F2 and F3 are similar in the "uh" and "o" sections indicating the vowels are similar (as Howell & Vause, 1986 reported).

Van Riper's argument could be applied to consonants that are prolonged. Prolonged /s/s. sound canonically like /s/. However there are different forms of /s/ that sound different which depend on what vowel follows. So, for example, an /s/ before an /i/ vowel sounds clearly different to an /s/ before an /u/ vowel. A possible explanation of /s/-prolongation could be that the wrong form of /s/ was selected and produced and when the speaker detected this, the transition to the following vowel could not be made

leading to the speaker prolonging the /s/. This can be tested by seeing whether the /s/ in words that have an /i/ following is acoustically identical when prolonged compared with when it is spoken fluently (in the same way this was done when comparing the vowels in a sequence of part word repetitions to the intended vowel at fluent release in the previous study). Speaker 61 at age 14 years 8 months produced two prolonged /s/s before an /i/ vowel – one at the beginning of the word "CD" at 47 s and one at the beginning of "CCF" at 115 s. Though there are no fluent /s/s before /i/ in this recording, a recording was made a month later (14 years 9 months) in which the speaker says 'CD' twice (both times fluently) (these appear at 112.2 and 116 s in this file). Oscillograms, spectrograms and cross sections were taken of the fluent and dysfluent /s/s. The main feature is that the spectra peak at around 5kHz and this applies to fluent and dysfluent forms of /s/ before /i/. Based on the acoustic similarity and informal listening to these examples, speakers appears to be articulating the form of /s/ that would permit coarticulation with the intended vowel that follows. Thus, an account of prolongation based on selection of the inappropriate allophone of /s/ does not seem correct.

So far only continuant phones (vowels and fricatives) have been discussed. These are produced with articulatory positions that do not change over time. It is possible that speakers who stammer have problems in controlling speech timing which would be reflected in phones that require changes in articulation over the time of their production. Plosive stop consonants are one class of sounds where such timing problems might be manifest. Plosives start with a short period of broad band energy that marks sound onset (the burst). The plosives can be divided into voiced (/b, d, g/) and voiceless (/p, t, k/) forms where corresponding pairs (e.g. /b/and /p/) have the same place of articulation. The differences in voicing arise because speakers control the timing of articulatory gestures in distinct ways for these two classes of plosives. After burst onset, vocal fold vibration starts almost immediately for voiced plosive (voicing gives rise to the pitch epoch markers mentioned in the pitch synchronous analysis part of section 3 of Appendix D which appear as striations in broad band spectrograms). In contrast, voiceless plosives have a period, after the burst, during which the phone is aspirated before voicing starts. The time between burst onset and onset of voicing can be used as a simple measure of voice onset time (VOT) that characterizes the difference between voiced (short VOT) and voiceless (long VOT) plosives. If a speaker who stammers has problems initiating voicing in time, this would be reflected in longer VOTs for /d/s than /t/s and make /d/s sound something like /t/s.

Some speakers who stammer appear to have problems initiating voicing so this should be reflected in the VOT measure. Speaker 1100 shows this characteristic. For instance, 367 s into his audio file he shows multi part word syllable repetitions on the word 'David' (prior to 'Hockney') which sound devoiced (i.e. the /d/s sounds like /t/s). A /d/ realized as /t/ should have a longer VOT (close to a voiceless plosives) than that of a true /d/. Acoustic analysis supports this notion, as you will see if you measure the VOT of the /d/s in the part word repetitions of the attempts at 'David' (prior to 'Hockney') and compare them with the VOT of /d/ in a fluent word (e.g. the "don't" that occurs at 80.7 s). You should also listen to the voiced sounds to confirm that they appear to be devoiced (i.e. the /d/ sounds like /t/).

Performance of HMM automatic dysfluency recognizer. Like fluent speech corpora, the data may also provide material for training automatic speech recognizers (Howell, Hamilton & Kyriacopoulos, 1986; Howell, Sackin & Glenn, 1997a, b; Noth, Niemann, Haderlein, Decher, Eysholdt, Rosanowski, & Wittenberg, submitted). Some hidden Markov model (HMM) utilities that have been used to construct a dysfluency recognizer are described in Appendix E. Performance of this recognizer on 5s intervals that experts agreed about for six test samples (0030_17y9m.1, 0061_14y8m.1, 0078_16y5m.1, 0095_7y7m.1, 0098_10y6m.1, 0138_13y3m.1, 0210_11y3m.1, 0234_9y9m.1) correctly classified 60% of intervals as stuttered/fluent. Though above chance, this is not particularly impressive. It does set a benchmark against which better dysfluency recognizers can be developed and assessed. One obvious improvement would be to be more selective when training the phone models (in the benchmark version, these were obtained from fluent speakers, not speakers who stammer). In addition to providing training material specifically for automating dysfluency counts, these data could be used more generally to test the robustness of recognizers developed for fluent speech. As it is claimed that recognizers perform at high levels when material is fluent, to what extent do failures in these algorithms coincide with stuttered dysfluencies? These are just some of the topics that the group at UCL are addressing and doubtless there are numerous other topics that the data and software will be helpful in investigating.

The JSRU coding scheme is presented in Appendix B. Modifications for the application of this scheme for dealing with stammered speech and codes for properties that are important for studying stammering are given in section 6 of Appendix C. The material that has been prepared according to this scheme can be used for analyses similar to those described in the studies cited above. A tutorial on manipulating transcriptions in SFS (with a focus on aligning them against the audio waveforms) is given in Appendix C. As stated earlier, SFS may prove particularly useful when researchers want to

visualize how disparate sources of information (e.g. duration and characterizations of phonetic difficulty) on different segmental units interact and lead to dysfluency. The aligned display convention should also prove useful when comparing the results of different analysis methods (such as those used for syntactic characterization of these materials) applied to the same stretch of data. A second SFS tutorial (Appendix D) covers basic aspects to do with acoustic analysis of speech that have been employed in some of the studies cited at the start of this section. Appendix E describes tools that could be used for developng HMM recognizers for dysfluent speech.

Acknowledgement

This research was supported by the Wellcome Trust.

References

Au-Yeung, J., Vallejo Gomez, I., & Howell, P. (2003). Exchange of disfluency from function words to content words with age in Spanish speakers who stutter. *Journal of Speech, Language and Hearing Research*, **46**, 754-765.

Dworzynski, K., & Howell, P. (2004). Predicting stuttering from phonetic complexity in German. *Journal of Fluency Disorders*, **29**, 149-173.

Dworzynski, K., Howell, P., Au-Yeung, J., & Rommel, D. (2004). Stuttering on function and content words across age groups of German speakers who stutter. *Journal of Multilingual Communication Disorders*, **2**, 81-101

Dworzynski, K., Howell, P., & Natke, U. (2003). Predicting stuttering from linguistic factors for German speakers in two age groups. *Journal of Fluency Disorders*, **28**, 95-113.

Howell, P. (1995). The acoustic properties of stuttered speech. In *Proceedings of the First World Congress on Fluency Disorders, Vol II*. Pp. 48-50. C. W. Starkweather & H. F. M. Peters (Eds). Nijmegen: Nijmegen University Press.

Howell, P. (2002). The EXPLAN theory of fluency control applied to the treatment of stuttering by altered feedback and operant procedures. In *Pathology and therapy of speech disorders*. Pp. 95-118. E. Fava (Ed.). Amsterdam: John Benjamins.

Howell, P. (2004). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communicative Sciences and Disorders*, **39**, 122-139.

Howell, P. (in press). Comparison of two ways of defining phonological words for assessing stuttering pattern changes with age in Spanish speakers who stutter. *Journal of Multilingual Communication Disorders*.

Howell, P., & Au-Yeung, J. (1995a). The association between stuttering, Brown's factors and phonological categories in child stutterers ranging in age between 2 and 12 years. *Journal of Fluency Disorders,* **20**, 331-344.

Howell, P., & Au-Yeung, J. (1995b). Syntactic determinants of stuttering in the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders*, **20**, 317-330.

Howell, P. & Au-Yeung, J. (2002). The EXPLAN theory of fluency control and the diagnosis of stuttering. In *Pathology and therapy of speech disorders*. Pp. 75-94. E. Fava (Ed.). Amsterdam: John Benjamins.

Howell, P., Au-Yeung, J., & Pilgrim, L. (1999). Utterance rate and linguistic properties as determinants of speech dysfluency in children who stutter: *Journal of the Acoustical Society of America,* **105**, 481-490.

Howell, P., Au-Yeung, J., & Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language and Hearing Research*, **42**, 345-354.

Howell, P., Au-Yeung, J., & Sackin. S. (2000). Internal structure of content words leading to lifespan differences in phonological difficulty in stuttering. *Journal of Fluency Disorders*, **25**, 1-20.

Howell, P., Hamilton, A., & Kyriacopoulos, A. (1986). Automatic detection of repetitions and prolongations in stuttered speech. In *Speech Input/Output: Techniques and Applications*. Pp. 252-256. London: IEE Publications.

Howell, P. Kapoor, A., & Rustin L. (1997). The effects of formal and casual interview styles on stuttering incidence. In *Speech Production: Motor Control, Brain Research and Fluency Disorders.* Pp. 515-520. W. Hulstijn, H. F. M. Peters & P. H. H. M. van Lieshout (Eds.). Amsterdam: Elsevier.

Howell, P., & Sackin, S. (2000). Speech rate manipulation and its effects on fluency reversal in children who stutter. *Journal of Developmental and Physical Disabilities*, **12**, 291-315.

Howell, P., Sackin, S., & Glenn, K. (1997a). Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: I. Psychometric procedures appropriate for selection of training material for lexical dysfluency classifiers. *Journal of Speech, Language and Hearing Research*, **40**, 1073-1084

Howell, P., Sackin, S., & Glenn, K. (1997b). Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers *Journal of Speech, Language and Hearing Research*, **40**, 1085-1096.

Howell, P., & Vause, L. (1986). Acoustic analysis and perception of vowels in stuttered speech. *Journal of the Acoustical Society of America*, **79**, 1571-1579.

Howell, P. & Williams, M. (1988). The contribution of the excitatory source to the perception of neutral vowels in stuttered speech. *Journal of the Acoustical Society of America*, **84**, 80-89.

Howell, P., & Williams, M. (1992). Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech. *Journal of the Acoustical Society of America*, **91**, 1697-1706.

Howell, P., & Young, K. (1990). Analysis of periodic and aperiodic components during fluent and dysfluent phases of child and adult stutterers' speech. *Phonetica*, **47**, 238-243.

Kadi-Hanifi, K., & Howell, P. (1992). Syntactic analysis of the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders*, **17**, 151-170.

Kully, D., & Boberg, E. (1988). An investigation of interclinic agreement in the identification of fluent and stuttered syllables. *Journal of Fluency Disorders*, **13**, 309-318.

Ladefoged, P. (1975). *A course in Phonetics*. New York: Harcourt, Brace, Jovanovich.

MacWhinney, B. (1995). *The CHILDES project*. Hillsdale NJ: Lawrence Erlbaum.

Noth, E., Niemann, H., Haderlein, T., Decher, M., Eysholdt, U., Rosanowski, F., & Wittenberg, T. (submitted). Automatic stuttering recognition using Hidden Markov models.

Rosen, S., & Howell, P. (1991). *Signals and Systems for Speech and Hearing*. London and San Diego: Academic Press.

Appendix A – Data description

This Appendix contains an indication where UCL's archive of recordings of speech from speakers who stutter are located and how they can be accessed.

1. Ancillary information about speaker and audio files

Some information we have about the speakers is given in the ACCESS file 'information table'. People who are familiar with ACCESS can use this to search and select the file for the recordings they want (by gender, age etc.). Each row of the table corresponds with one of the 138 files in the directory. The filename is given in the column labeled 'file id'. This starts with a code for gender (M/F), then has underscore and a four figure code to identify speaker, UCL group, underscore class number (as given in Table 1), underscore age NNvNm, followed by undersore and a number indicating which recording in that month this represents. The second column gives gender (M for male and F for female). The age of the speaker at the recording session is given in months in column three. The fourth column indicates where the recording was made (clinic, UCL or home). Recording conditions are indicated in column five as either as quiet room (QR) or sound-treated room (STR). The sixth column gives the type of therapy received (either including the family, F, or holistic, H). An indication is given about the time (in months) between the recordings and therapy in the seventh column. The eighth column indicates whether the speaker had any history of hearing problems. The ninth and tenth columns indicate whether the speaker had a history of language problems or special educational needs respectively, and the eleventh, twelfth and thirteenth columns indicate whether any manual transcripts are available (orthographic, phonetic and, for the files that have phonetic transcriptions, those which are time aligned, respectivly).

2. Formats available

For CHILDES. Transcriptions are prepared according to TEXTIN (MacWhinney, 1995) and corresponding audio files are also available in WAV format. The WAV files have been linked to the CHAT files. CLAN has options that allow PRAAT programs to process the associated WAV files. Other information that CHAT files include in their headers are available in the ACCESS file described above.

CHAT files + WAV = CHILDES

For PRAAT. The transcriptions have been converted to PRAAT TextGrids. PRAAT provides acoustic analysis facilities for dealing with WAV files that are also available in the directory.) The way PRAAT works is described on the PRAAT website (*http://www.praat.org*).

TextGrids + WAV = PRAAT

For SFS. There is an SFS file corresponding to each of the recordings. Most of these just contain the audio file. The remaining ones have phonetic transcriptions that have been aligned manually against the audio file. Separate orthographic and phonetic transcriptions are available for some of the files as text files. The phonetic transcriptions are in JSRU format (see Appendix B), but it should be easy to translate them to other phonetic formats. The audio waveforms and aligned transcriptions can be displayed and manipulated using the whole rang of SFS utilities. The separate orthographic and phonetic files can be used in the transcription exercises indicated in Appendix C.

Other format. All data are also available as MP3 files.

The speech and ancillary data can be accessed at:

http://www.uclass.psychol.ucl.ac.uk

3. Description of discs

The dataare on eight CDs that follow the directory structure of the online Release 1 (August 2004) dataset.

Disc 1 contains all the available transcription data. It also contains an MS Access database and an HTML formatted page of the files in Release 1: flat orthographic and phonetic transcriptions and time aligned transcriptions in SFS Annotation, CHILDES CHAT and Praat TextGrid format.

Disc 2 contains all audio data in compressed mp3 format.

Discs 3-5 contain all audio data in SFS format and Disc 3 also contains the available SFS audio data with time aligned annotations and also those files used in the worked examples.

Discs 6-8 contain all the audio data in wav format.

Appendix B The JSRU alphabet

This Appendix gives the Joint Speech Research Unit (JSRU) alphabet.

1. The JSRU Alphabet

The JSRU (Joint Speech Research Unit) alphabet is a phonetic transcription system that was developed for text-to-speech synthesis. It has a machine readable format. The correspondence between 'qwerty' keyboard characters and phonetic symbols is given in Table B.1 for consonants and Table B.2 for vowels.

Table B.1. JSRU symbols and IPA equivalents for English consonants. IPA symbols are only given when they differ from the JSRU symbols.

JSRU	**IPA**	**Sound info**	**Example** (word initial)	**Example** (non-initial)
p	(same)	p sound	pen	Lip
b	(same)	b sound	bad	nib
t	(same)	t sound	tea	Light
d	(same)	d sound	do	Lad
k	(same)	k sound	cat	crack
g	(same)	g sound	got	Fog
ch		ch sound	chin	watch
j		j sound	June	village
f	(same)	f sound	food	off
v	(same)	v sound	voice	give
th		th sound	thin	Tenth
dh		dh sound	then	with
s	(same)	s sound	same	success
z	(same)	z sound	zoo	was
sh		sh sound	show	wash
zh		zh sound	genre	beige
h	(same)	h sound	happy	behave
m	(same)	m sound	man	swim
n	(same)	n sound	know	gone
ng		ng sound	N/A	sing
l	(same)	l sound	leg	girl
r		r sound	red	arrow
y		y sound	year	value
w	(same)	w sound	wet	quick
x		x sound	N/A	Loch (Scots)
gx		gx sound	got in Cockney	glottal stop

Table B.2. JSRU symbols and IPA equivalents for English vowels. IPA symbols are only given when they differ from the JSRU symbols.

JSRU	IPA	Sound info	Example
i	ɪ	i sound	sit
o	ɒ	o sound	got
oo	ʊ	oo sound	put
aa	æ	aa sound	hat
e	e	e sound	ten
U	ʌ	u sound	cup
a	ə	a sound	ago
ee	iː	ee/ey sound	see
aw	ɔː	aw sound	saw
uu	uː	uu sound	too
ar	ɑː	ar sound	arm
er	ɜː	er sound	fur
ai	eɪ	ai sound	page
ie	aɪ	ie sound	eye
oi	ɔɪ	oi sound	boy
oa	əʊ	oa sound	home
ou	aʊ	ou sound	now
ia	ɪə	ia sound	beer
ei	eə	ei sound	bare
ur	ʊə	ur sound	tour

Appendix C Orthographic and Phonetic Annotation with SFS

*The previous Appendix described the JSRU alphabet but did not indicate how the transcriptions can be linked with the audio signal. SFS provides an environment for representing audio and transcriptional information together. This Appendix provides a tutorial introduction to the use of SFS for the orthographic and phonetic transcription of a speech recording, including tools for automatic alignment of phonetic transcriptions to the signal. Some software is presented and described in this Appendix but programming experience is not assumed. You may not need to do all steps described in this Appendix. For instance, if you only intend to use the data supplied in Appendix A, the sections on acquiring the audio signal are not needed. This tutorial refers to versions 4.6 and later of SFS and appears in the documentation on the SFS website. Visit the SFS website to obtain your software (*http://www.phon.ucl.ac.uk/resource/sfs/).

1. Acquiring and Chunking the audio signal

Acquiring the signal

You can use the SFSWin program to record directly from the audio input signal on your computer. Only do this if you know that your audio input is of good quality, since many PCs have rather poor quality audio inputs. In particular, microphone inputs on PCs are commonly very noisy.

To acquire a signal using SFSWin, choose File|New, then Item|Record. See Figure C.1.1. Choose a suitable sampling rate, at least 16000 samples/sec is recommended. It is usually not necessary to choose a rate faster than 22050 samples/sec for speech signals.

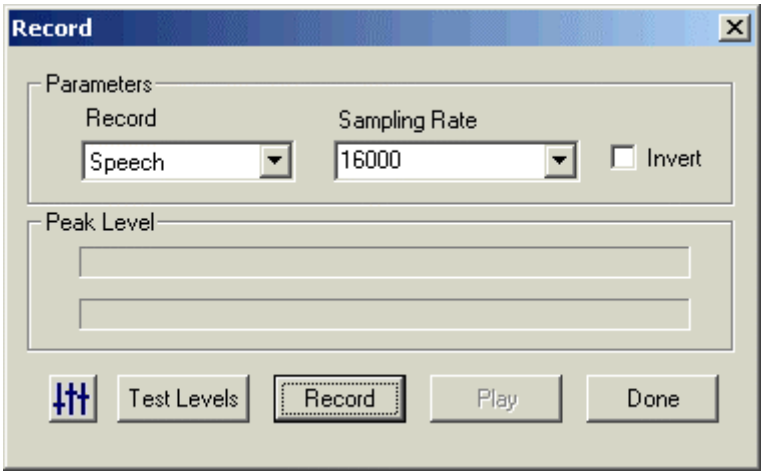


Figure C.1.1 - SFSWin record dialog

If you choose to acquire your recording into a file using some other program, or if it is already in an audio file, choose Item|Import|Speech rather than Item|Record to load the recording into SFS. If the file is recorded in plain PCM format in a WAV audio file, you can also just open the file with File|Open. In this latter case, you will be offered a choice to "Copy contents" or "Link to file" to the WAV file. See Figure C.1.2. If you choose copy, then the contents of the audio recording are copied into the SFS file. If you choose link, then the SFS file simply "points" to the WAV file so that it may be processed by SFS programs, but it is not copied (this means that if the WAV file is deleted or moved SFS will report an error).

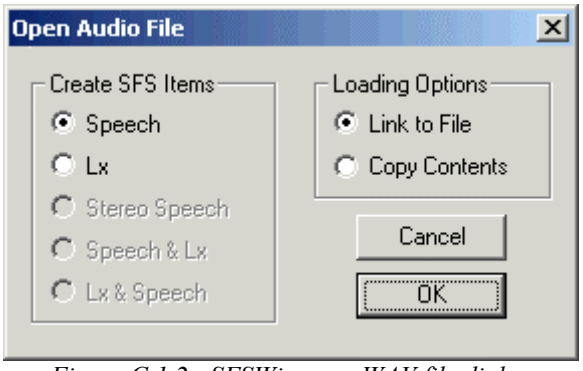


Figure C.1.2 - SFSWin open WAV file dialog

Preparing the signal

If the audio recording has significant amounts of background noise, you may like to try to clean the recording using Tools|Speech|Process|Signal enhancement. See Figure C.1.3. The default setting is "100% spectral subtraction"; this subtracts 100% of the quietest spectral slice from every frame. This is a fairly conservative level of enhancement, and you can try values greater than 100% to get a more aggressive enhancement, but at the risk of introducing artefacts.

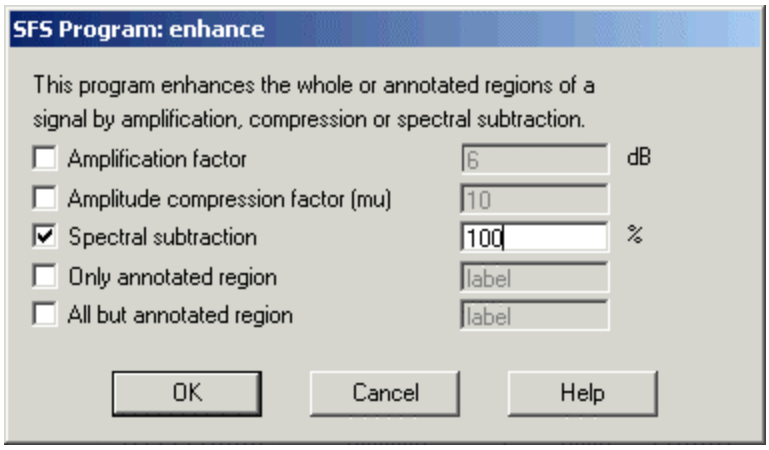


Figure C.1.3 - SFSWin enhancement dialog

It is also suggested at this stage that you standardise the level of the recording. You can do this with Tools|Speech|Process|Waveform preparation, choosing the option "Automatic gain control (16-bit)" (AGC). No noise reduction (spectral subtraction) nor AGC have been made on the recordings that can be accessed using the information in Appendix A.

Chunking the signal

If your audio recording is longer than a single sentence, you will almost certainly gain from first chunking the signal into regions of about one sentence in length. Chunking involves adding a set of annotations which delimit sections of the signal. The advantages of chunking include:

- it means that transcription is roughly aligned to the signal .
- it makes it easier to navigate around the signal.
- it improves the performance of automatic phonetic alignment.
- it allows the export of a "click-to-listen" web page using VoiScript (see below).
- it allows us to use SFS annotations to store the transcription, since SFS limits annotations to 250 characters.

An easy way to chunk the signal is to automatically detect pauses using the "npoint" program. This takes a speech signal as input and creates a set of annotations which mark the beginning and end of each region where someone is speaking. It is a simple and robust procedure based on energy in the signal. To use this, select the speech item and choose Tools|Speech|Annotate|Find multiple endpoints. See Figure C.1.4. If you know the number of spoken chunks in the file (it may be a recording of a list of words, for example), enter the number using the "Number of utterances to find" option, otherwise choose the "Auto count utterances" option. Put "chunk" (or similar) as the label stem for annotation.

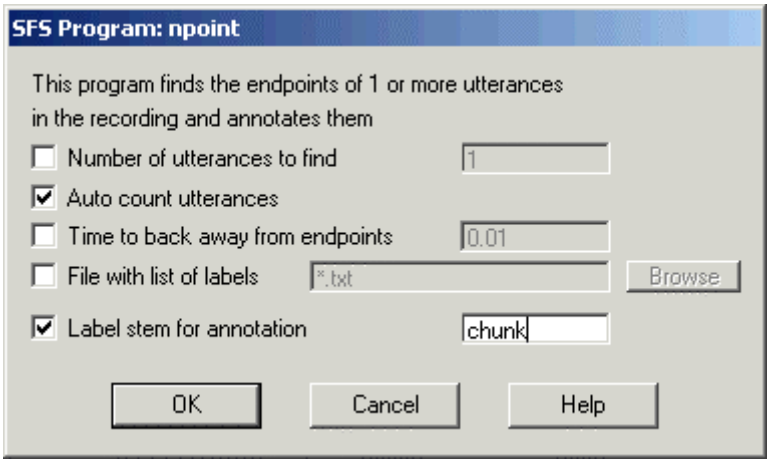


Figure C.1.4 - SFSWin find multiple endpoints dialog

If you view the results of the chunking you will see that each spoken region has been labeled with "chunk*dd*", while the pauses are labelled with "/". See Figure C.1.5 for the results of applying this procedure to one of the files provided.

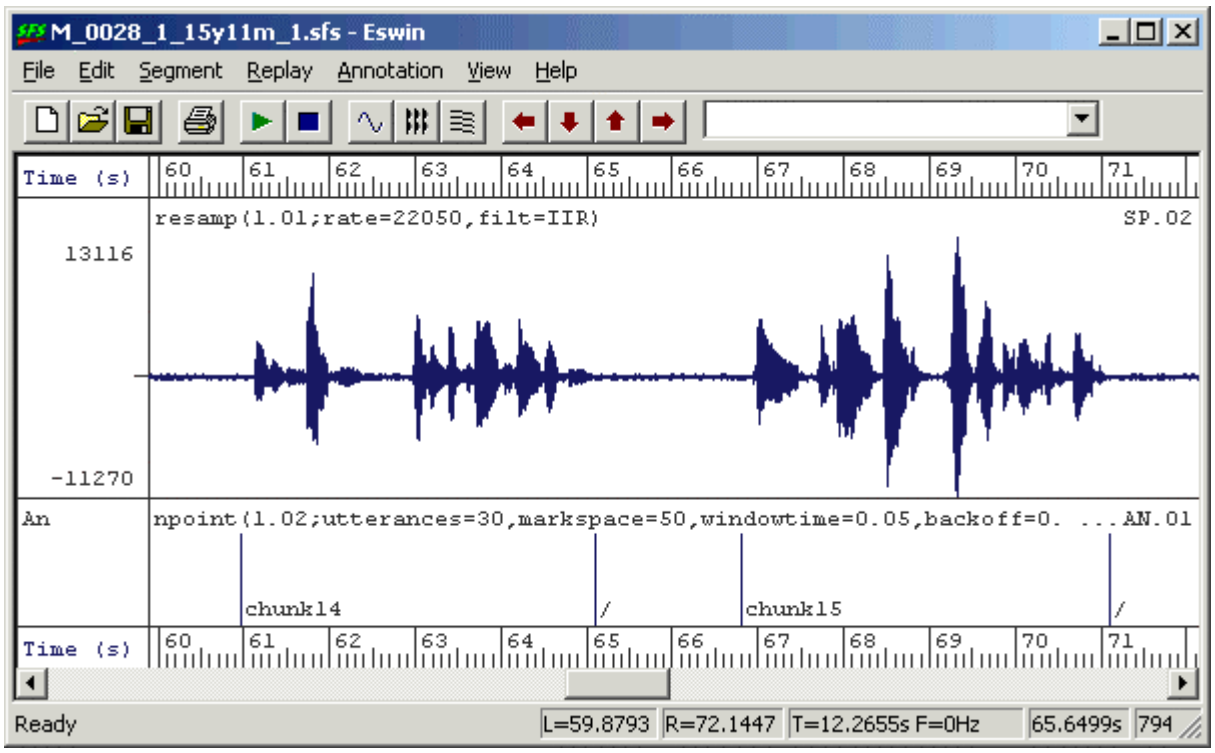


Figure C.1.5 - chunked signal

If the chunking has not worked properly, or if you want to chunk the signal by hand, you can use the manual annotation facility in Eswin. To do this, select the signal you want to annotate and choose Item|Display to start the eswin program. Then choose eswin menu option Annotation|Create/Edit Annotations, and enter either a set name of "chunks" to create a new set of annotations, or enter "endpoints" to edit the set of annotations produced by npoint.

When eswin is ready to edit annotations you will see a new region at the bottom of the screen where your annotations will appear. To add a new annotation, position the left cursor at the time where you want the annotation to appear. Then type in the annotation into the annotation box on the toolbar and press [RETURN]. The annotation should appear at the position of the cursor. See Figure C.1.6.

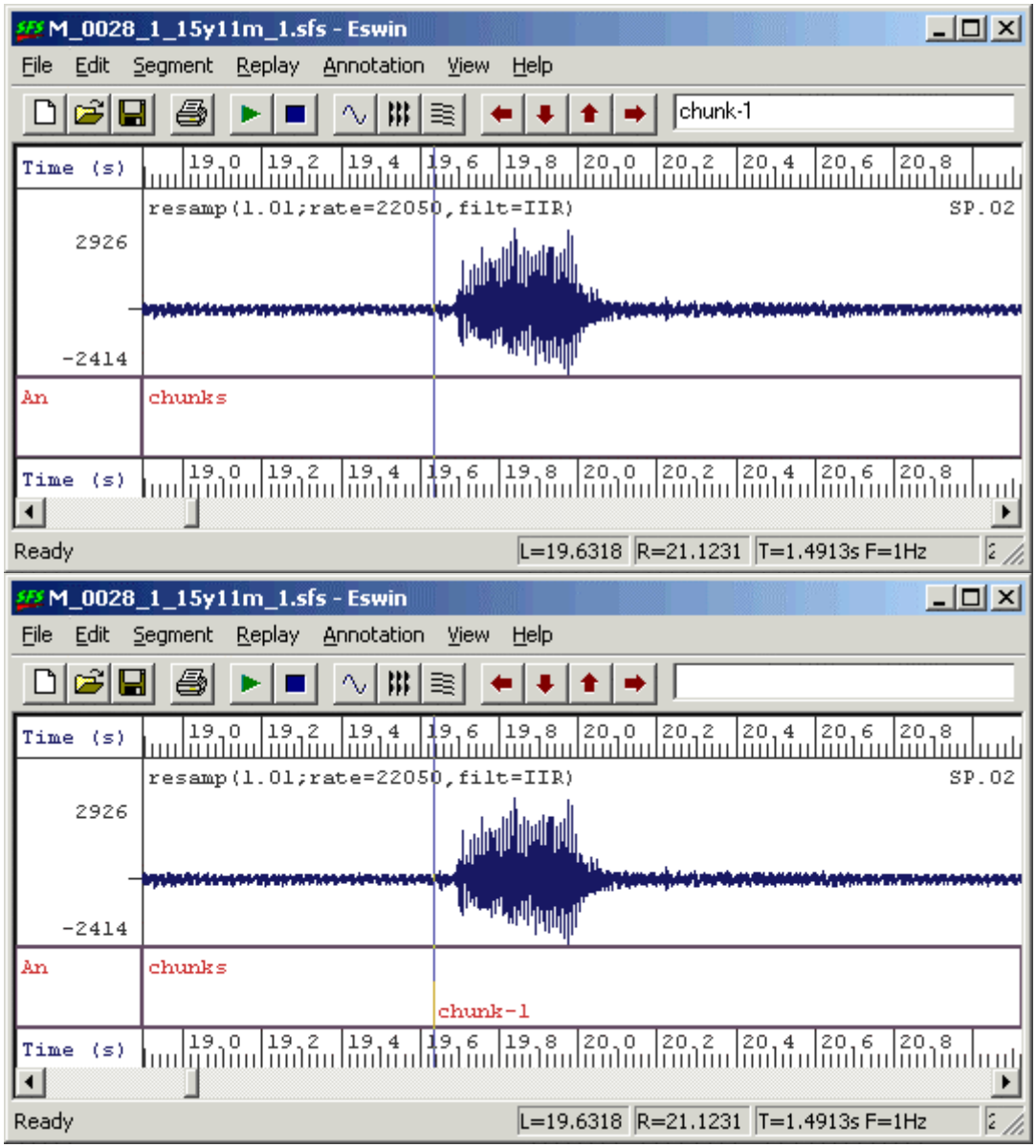


Figure C.1.6 - adding an annotation in eswin

To move an annotation with the mouse, position the mouse cursor on the annotation line within the bottom annotation box. You will see that the mouse cursor changes shape into a double-headed arrow. Press the left mouse button and drag the annotation left or right to its new location. This is also an easy way to correct chunk endpoints found automatically by npoint.

Finally, to hear if the chunking has worked properly, you can listen to the chunked recording using the SFS wordplay program. This program is not on the SFSWin menus, so to run it, choose Tools|Run program/script then enter "wordplay -SB" in the "Program/script name" box. See figure C.1.7. This will replay each chunk in turn, separating the chunks with a small beep.

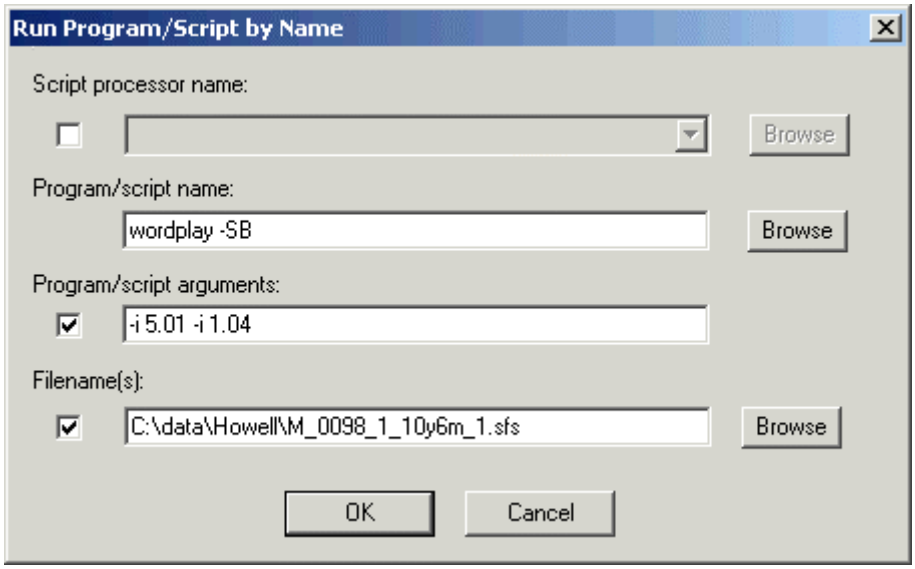


Figure C.1.7 - SFSWin run program dialog

2. Orthographic transcription

Entering orthographic transcription

Assuming that your recording has been chunked into sentence-sized regions, the process of orthographic transcription is now just the process of replacing the "chunk" labels with the real spoken text. The result will be a new annotation item in the file, but where each annotation contains the orthographic transcription of a chunk of signal. See Figure C.2.1.

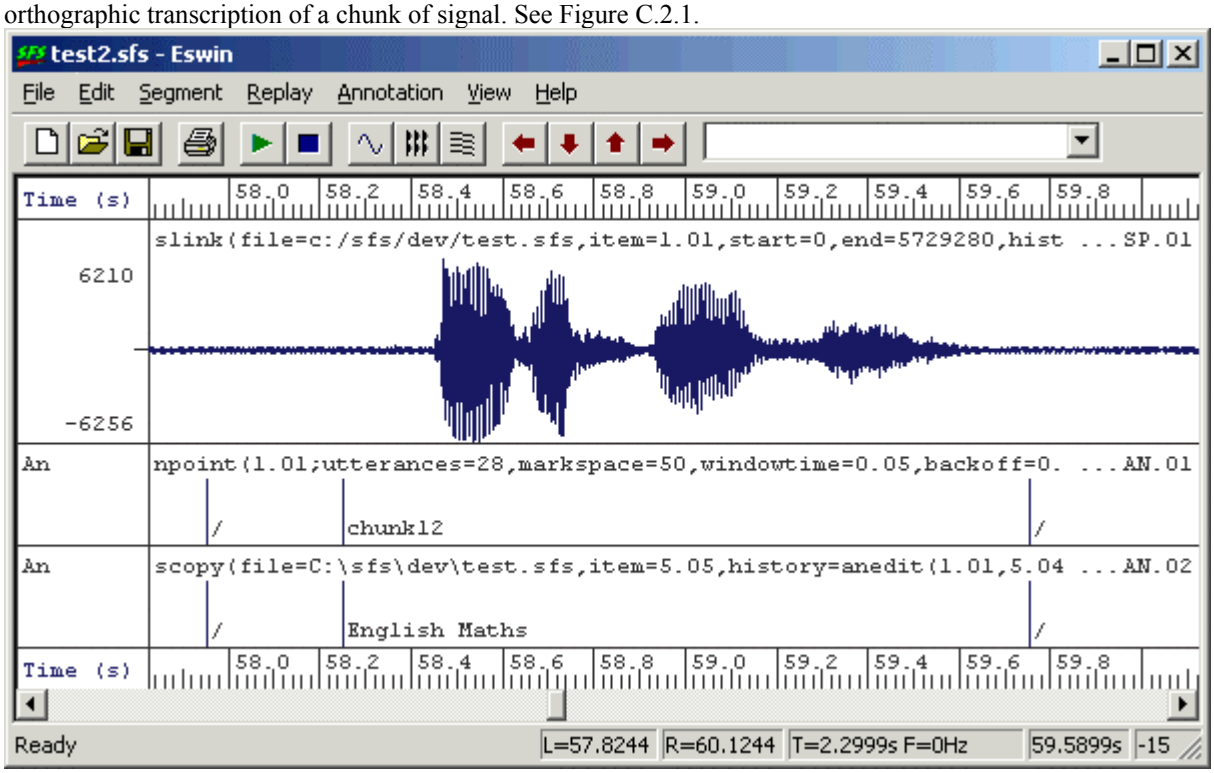


Figure C.2.1 - Speech chunks and orthographic transcription

You *can* edit annotation labels using the eswin display program, but it is not very easy - you have to overwrite each annotation label with the transcription. A much easier way is to use the anedit annotation label editor program. This program allows you to listen to the individual annotated regions and to edit the labels of annotations without affecting their timing. To run anedit, select a speech item and the annotation item containing the chunks and choose Tools|Annotations|Edit Labels. Since you

are mapping one set of annotations into another, change the "output" annotation type to "orthography". See Figure C.2.2.

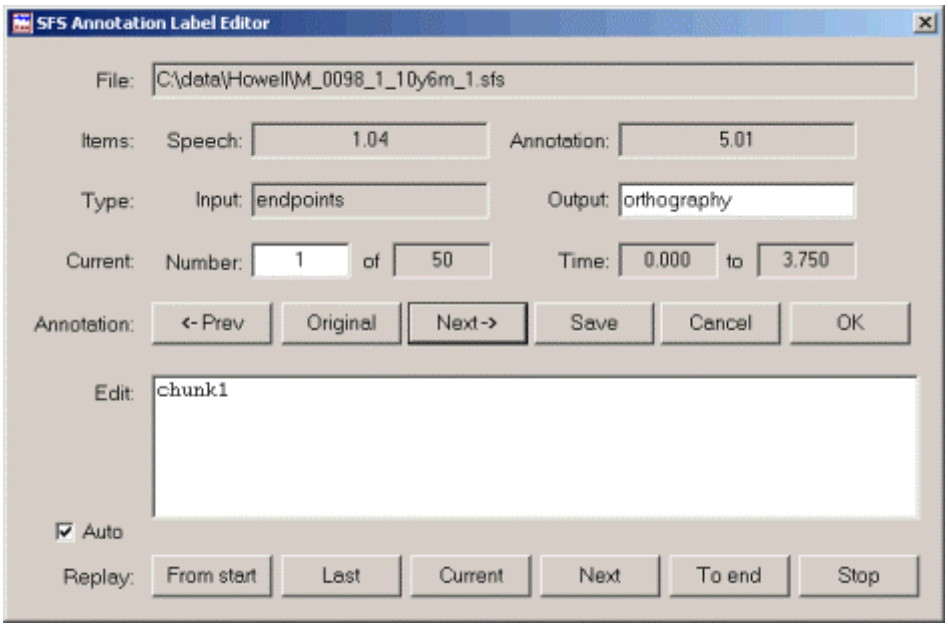


Figure 2.2 - anedit window

The row of buttons in the middle of the annotation editor window control the set of annotations:

<–Prev
Move to previous annotation in the list.
Original
Reload the original annotation label at this position.
Next–>
Move to next annotation in the list.
Save
Save the changes you've made so far to the SFS file.
Cancel
Forget annotation label changes and exit.
OK
Save annotation label changes and exit.

The row of buttons at the bottom of the annotation editor window control the replay of the speech signal:

From start
Replay from the start of the file to end of the current annotated region.
Last
Replay from the start of the last annotated region to the end of the current annotated region.
Current
Replay the current annotated region.
Next
Replay from the start of the current annotated region to the end of the next annotated region.
To end
Replay from the start of the current annotated region to the end of the file.
Stop
Stop any current replay.

The "Auto" replay feature causes the current annotated region to be replayed each time you change to a different annotation.

To use anedit for entering orthographic transcription, first check that the "Auto" replay feature is enabled and that you are positioned at the first chunk of speech. Replay this with the "Current" button, select and over-write the old label with the text that was spoken. Then press the [RETURN] key. Two things should happen: first you should move on to the next chunk in the file and second that chunk of signal should be replayed. You can now proceed through the file, entering a text transcription and pressing [RETURN] to move on to the next chunk. If you need to hear the signal again, use the buttons

at the bottom of the screen. It is suggested that every so often you save your transcription back to the file with the "Save" button. This ensures you will not lose a lot of work should something go wrong.

One **word of warning**: at present SFS is limited to annotations that are less than 250 characters long. Anedit prevents you from entering longer labels. There is no limit to the *number* of labels however.

Conventions

It is worth thinking about some conventions for entering transcription. For example, should you start utterances with capital letters, or terminate them with full stops? Should you use punctuation? Should you use abbreviations and digits? Should you mark non-speech sounds like breath sounds, lip smacks or coughs?

Here is one convention that you might follow, which has the advantage that it is also maximally compatible with SFS tools.

- put all words except proper nouns in lower case.
- do not include any punctuation.
- spell out all abbreviations and numbers, i.e. "g. c. s. e." not "GCSE", "one hundred and two" not "102", "ten thirty" not "10:30".
- mark non-speech sounds in a special way, e.g. "[cough]".

For pause regions you can either choose to label these using a special symbol of your own (e.g. "[pause]"), or leave them annotated as "/", or label them with the SAMPA symbol for pause which is "...".

Making a clickable script

Once you have a chunked and transcribed recording you can distribute your transcription as a "clickable script" using the VoiScript program (available for free download from **http://www.phon.ucl.ac.uk/resource/voiscript/**). The VoiScript program will display your transcription and replay parts of it in response to mouse clicks on the transcription itself. This makes it a very convenient vehicle for others to listen to your recording and study your transcription.

VoiScript takes as input a WAV file of the audio recording and an HTML file containing the transcription coded as links to parts of the audio. Technical details can be found on the VoiScript web site. To save your recording as a WAV file, choose Tools|Speech|Export|Make WAV file, and enter a suitable folder and name for the file. The following SML script can be used to create a basic HTML file compatible with VoiScript:

```
/* anscript.sml - convert annotation item to VoiScript HTML file */

/* takes as input file.sfs and outputs HTML
            assuming audio is in file.wav */

main {
  string basename
  var    i,num

  i=index("\.",$filename);
  if (i) basename=$filename:1:I-1 else basename=$filename;

  print "<html><body><h1>",basename,"</h1>\n";

  num=numberof(".");
  for (I=1;i<=num;i=i+1) if (compare(matchn(".",i),"/")!=0) {
    print "<a name=chunk",i:1
    print " href='",basename,".wav#",timen(".",i):1:4
    print "#",(timen(".",i)+lengthn(".",i)):1:4,"'>"
    print matchn(".",i),"</a>\n"
  }

  print "</body></html>\n"
}
```

Copy and paste this script into a file "anscript.sml". Then select the annotation item you want to base the output on and choose Tools|Run SML script. Enter "anscript.sml" as the SML script filename and the name of the output HTML file as the output listing filename in the same directory as the WAV audio file.

If you now open the output HTML file within the VoiScript program, you will be able to read and replay parts of the transcription on demand, see Figure C.2.3.

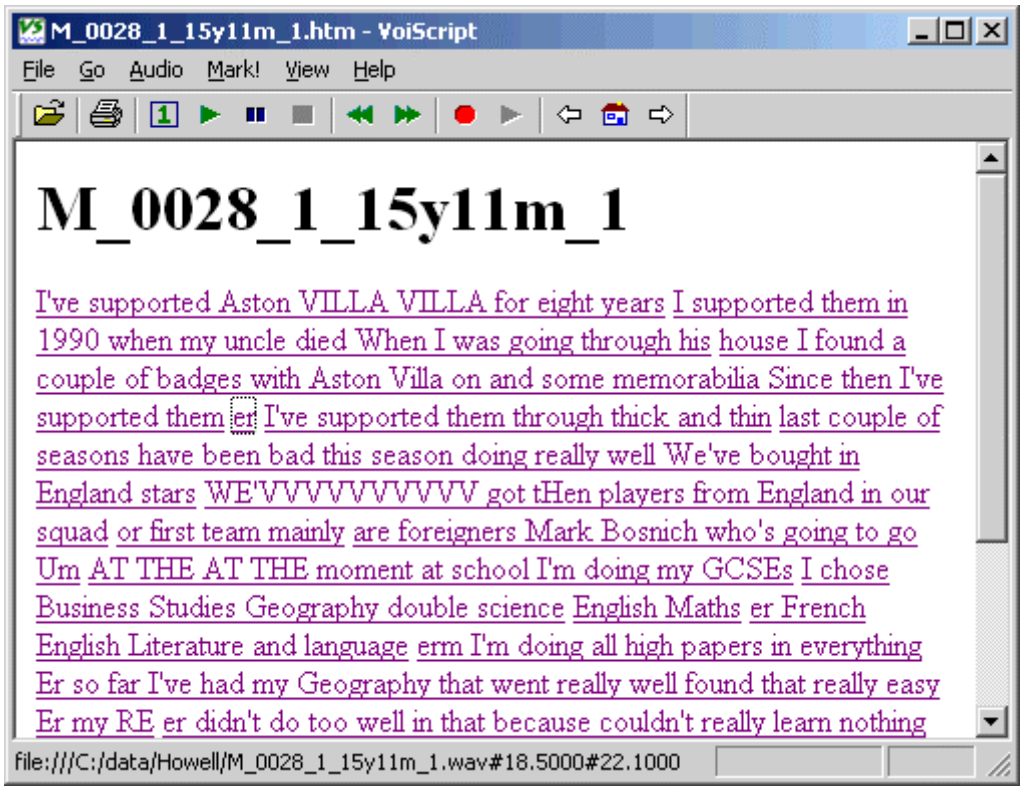


Figure C.2.3 - example VoiScript clickable script

3. Phonetic transcription

Spelling to sound

We now have a chunked orthographic transcription of our recording roughly aligned to the audio signal. The next stage is to translate the orthography for each chunk into a phonetic transcription. If we know the language, this is a largely mechanical procedure of looking up words in a dictionary. If the language is English, the mechanical part of the process can be performed by the antrans program.

The SFS program antrans performs the phonetic transcription of orthography using a built-in English pronunciation dictionary. The program takes orthographic annotations as input and produces transcribed annotations as output, in which only the content of the labels has been changed. See Figure C.3.1

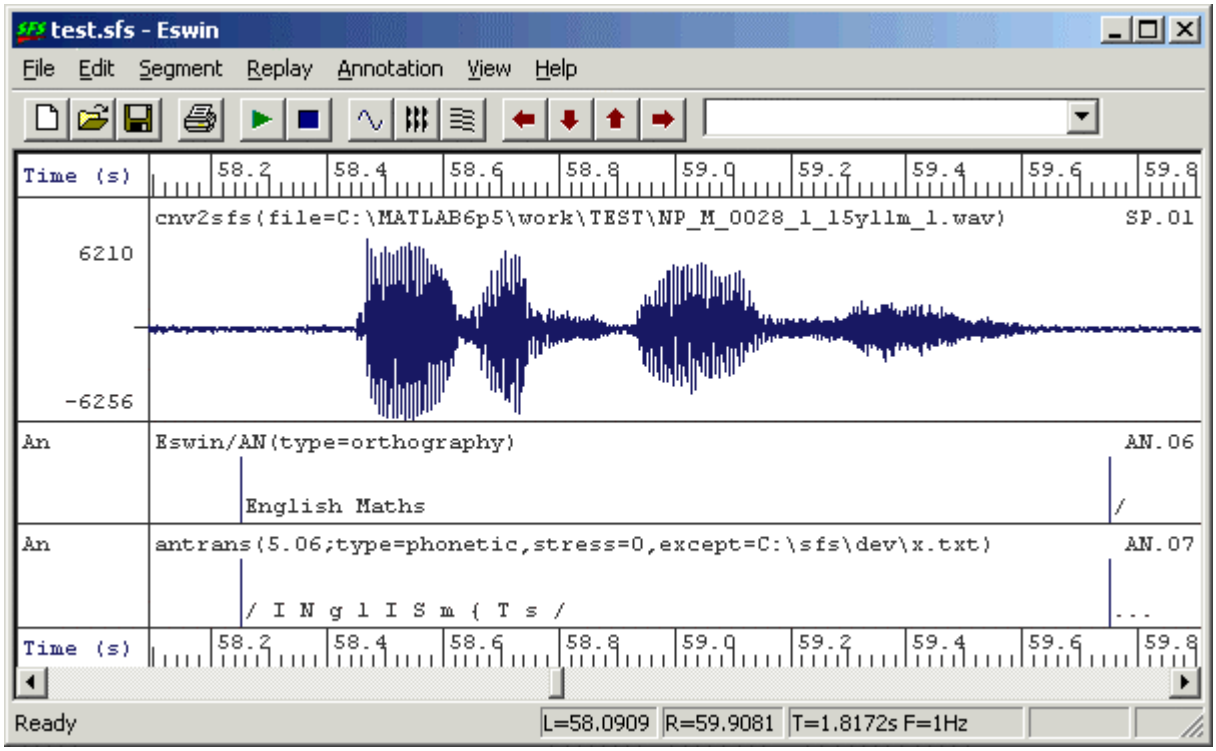


Figure C.3.1 - Transcribed annotations

It will almost certainly be the case that antrans will do an imperfect job in any real situation, since:

- it only produces a single pronunciation for each word, and that may not be the pronunciation used by the speaker;
- it may not know all the words used: although it has a large dictionary, it cannot know all names, places and abbreviations;
- it does not take into account any possible or actual contextual changes to pronunciation, such as assimilations and elisions;
- it can only guess that each chunk begins and ends in silence.

To run antrans, select the input annotation item and choose Tools|Annotation|Transcribe labels. The first time this is run, collect a list of words that antrans doesn't know by using the 'Missing word list' option, see Figure C.3.2. After the program has run, edit the word list (in Windows notepad for example) and add a transcription to each word, saving the resulting file as an exceptions list. This can then be incorporated in a second run of antrans (you can delete the output of the first run), see Figure C.3.3.

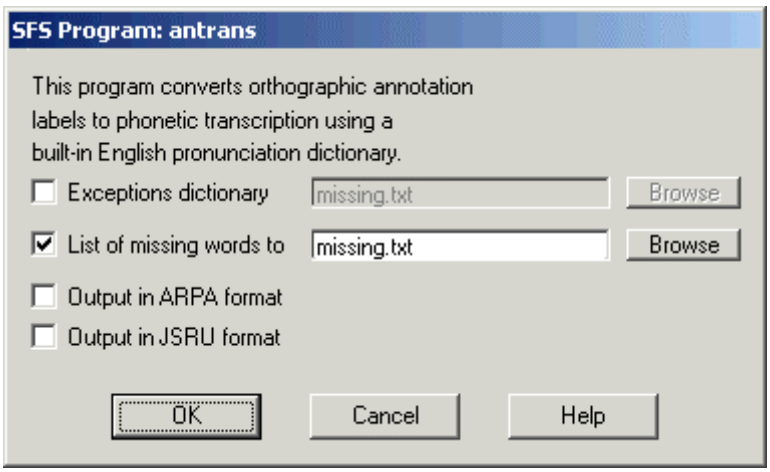


Figure C.3.2 - SFSWin Transcribe labels dialog (1)

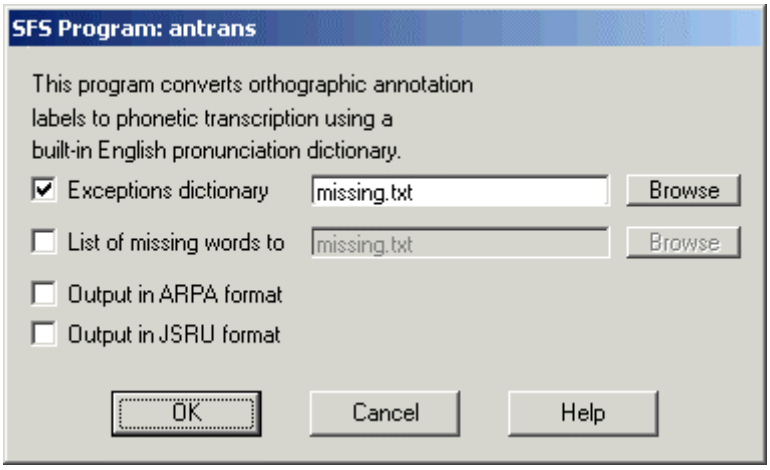


Figure C.3.3 - SFSWin Transcribe labels dialog (2)

The format of the exceptions file is as follows. It is a text file where each line is the pronunciation of a single word. A word is a sequence of printable characters that do not contain a space. The spelling of the word is followed by a TAB character, and then the transcription follows in SAMPA notation. It is usually not necessary to separate the SAMPA segment symbols with spaces, but it does not do any harm. Include stress symbols only if you intend to use them later. Here is an example:

```
1990    naInti:n naIntI
Bosnich bQznItS
MATHSSSSSSS     m{Ts
WE'VVVVVVVVV    wi:v
```

A simple way to correct the transcription is to use the anedit program again, just as we did for entering orthographic transcription in section C.2.

Transcription systems

The SFS tools are designed to work with the SAMPA transcription system by default, but antrans can also use transcriptions in ARPA and JSRU systems. The table below gives a comparison of the symbol system as compared to the IPA.

IPA Keyword SAMPA ARPA JSRU		IPA Keyword SAMPA ARPA JSRU	
P		ɪ	
Pin		pit	
P		I	
P		ih	
P		i	
B		e	
Bin		pet	
B		e	
B		eh	
B		e	
T		æ	
Tin		pat	
T		{	

T	ae
T	aa
D	ɒ
Din	pot
D	Q
D	oh
D	o
K	ʌ
Kin	cut
K	V
K	ah
K	u
G	ʊ
Give	put
G	U
G	uh
G	oo
tʃ	ə
chin	another
tS	@
ch	ax
ch	a
dʒ	iː
gin	ease
dZ	i:
jh	iy
j	ee
f	eɪ
fin	raise
f	eI
f	ey
f	ai
v	aɪ
vim	rise
v	aI
v	ay
v	ie
θ	ɔɪ
thin	noise
T	OI
Th	oy
Th	oi
Ð	uː

This	lose
D	u:
Dh	uw
Dh	uu
S	əʊ
Sin	nose
S	@U
S	ow
S	oa
Z	aʊ
Zing	rouse
Z	aU
Z	aw
Z	ou
ʃ	ɜː
shin	furs
S	3:
Sh	er
Sh	er
ʒ	ɑː
measure	stars
Z	A:
Zh	aa
Zh	ar
H	ɔː
Hit	cause
H	O:
H	ao
H	aw
M	ɪə
Mock	fears
M	I@
M	ia
M	ia
N	eə
Knock	stairs
N	e@
N	ea
N	ei
Ŋ	ʊə
Thing	cures
N	U@
Ng	ua
Ng	ur

R
Wrong
R
R
R

L
Long
L
L
L

W
Wasp
W
W
W

J
Yacht
J
Y
Y

ʔ
network
ʔ
ʔ
gx

x
loch
x
x
x

In addition, the following symbols are used to mark stress and silence:

IPA	Description	SAMPA	ARPA	JSRU
ˈ	*primary stress*	"		"
ˌ	*secondary stress*	%		'
	Silence	/	sil	q

4. Aligning phonetic transcription

At this point we have a chunked phonetic transcription: each spoken chunk of the signal is annotated with a unit of phonetic transcription. The next stage is to break up the transcription into individual segment labels and roughly align the labels to the signal. See Figure C.4.1. A basic level of alignment can be performed by the SFS analign program.

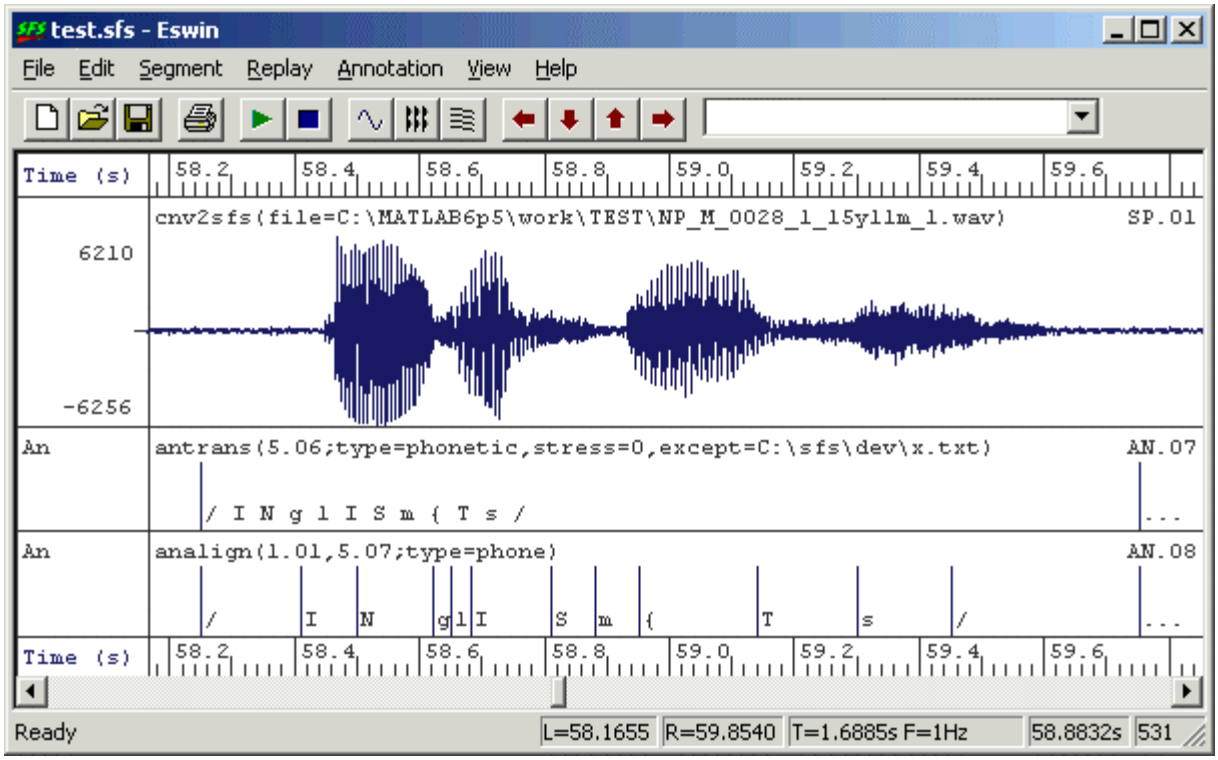


Figure C.4.1 - Chunked vs. aligned transcription

Automatic alignment

Analign has two modes of operation. In the first mode, input is a set of transcribed chunks in which the start and end points of the chunks are fixed. The program then finds an alignment between the segments in the transcription and the signal region identified by the chunk. In the second mode, the program chooses chunks on the basis of pause labels, and all phonetic annotations between the pauses are realigned. By default pauses are identified by labels containing the SAMPA pause symbol "...". You can use the first mode to get a basic alignment, then you can use the second mode to refine the alignment by adding or deleting phonetic annotations and re-running analign.

To align chunked phonetic transcription, select the input speech and annotation items and choose Tools|Annotation|Auto-align phone labels. Choose option "Fixed label boundaries" to only perform alignment *within* a label. See Figure C.4.2.

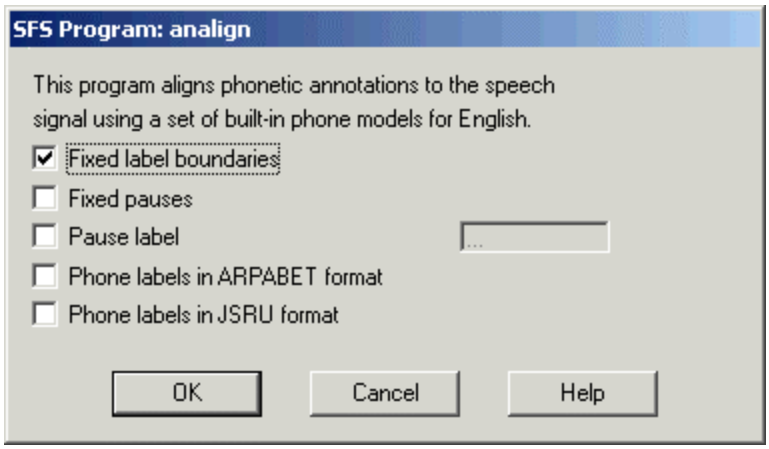


Figure C.4.2 - SFSWin Align Labels dialog (1)

The automatic alignment is performed using a set of phone hidden-Markov models which have been trained on Southern British English. You may need to replace these for other languages and accents. Look at the manual page of analign for details. The HMMs that come with SFS have been built using the Cambridge hidden Markov modelling toolkit HTK (also see Appendix E).

Automatic alignment is an approximate process, and you will almost certainly see places in the aligned transcription where the alignment is not satisfactory. Common kinds of problems are:

1. Segments stretched over unmarked pauses.
2. Segments compressed when smoothed or elided in rapid speech.
3. Poor alignment in consonant clusters and unstressed syllables in rapid speech.
4. Poor identification of speech-to-silence boundaries.
5. Poor alignment for syllable-initial glides and syllable-final nasals.

You can either correct the alignment manually or you can make changes to the transcription and run analign again. We'll describe these in turn.

Manual editing of transcription alignment

To edit a set of annotations, select a speech signal and the annotations to be edited and choose Item|Display. The waveform and the input annotations are displayed within SFS program eswin. Then right-click with the mouse in the box at the left of the annotations and choose menu option "Edit annotations". An editable copy of the set of annotations will then appear at the bottom of the screen.

Eswin has a number of special facilities to help in the correction of annotation alignments. To demonstrate these, zoom into a region of the signal so that individual annotations are clearly visible. Then click the left mouse button to display the vertical cursor. You will then find that:

- the **left and right arrow keys [<−] and [−>]** shift the left cursor one pixel to the left and right;
- pressing **[Ctrl] together with the arrow keys** will cause the left cursor to jump from one annotation to the previous/next annotation. The annotation label is also copied into the annotation edit box;
- with the left cursor on an annotation, pressing **[Shift] together with the arrow keys** will slide the annotation one pixel left and right;
- with the left cursor on an annotation, pressing the **[Delete] key** will delete an annotation.

You can also delete an annotation by deleting the contents of the annotation edit box and pressing [Return] while the cursor is positioned on an annotation. Figure C.4.3 shows an annotation being moved using the arrow keys.

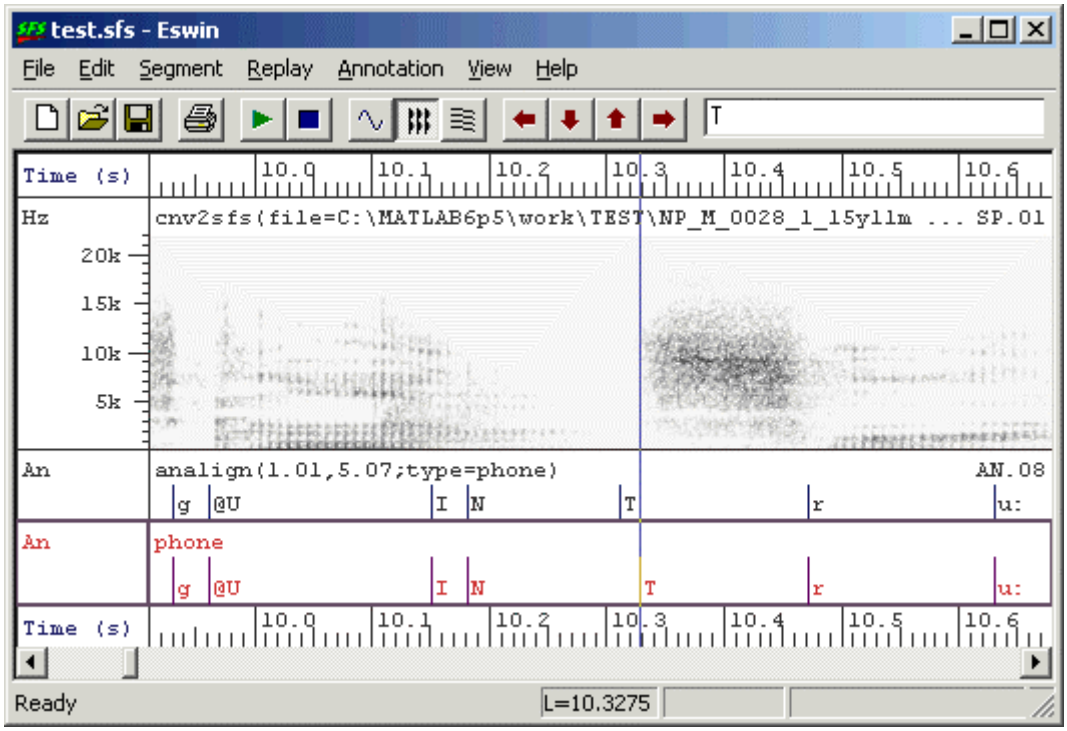


Figure C.4.3 - Manual editing of annotations

Semi-automatic alignment correction

If the automatic alignment has failed for fairly obvious reasons, it may be more efficient to redo the alignment with the problem fixed than to reposition every annotation manually. For example, a common problem is a failure to mark short pauses that occur within utterances. It is easy to add these pauses as new annotations (with "/" symbols) and to re-do the automatic alignment.

Because we have aligned the transcription once, we do not want analign to preserve the current annotation label boundaries. Instead we probably want to preserve the position of major pauses in the transcription (marked with "..." symbols). To re-do the alignment this way, select the speech signal and the edited aligned annotations and choose Tools|Annotations|Auto-align phone labels, but choosing option "Fixed pauses", see Figure C.4.4. If you use a different symbol to "..." for pauses, enter the symbol as the "Pause label" parameter.

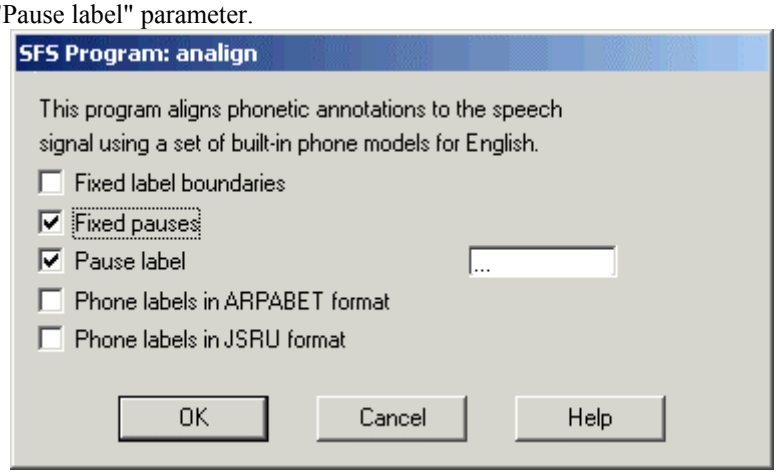


Figure C.4.4 - SFSWin Align Labels dialog (2)

5. Verification and Post-processing

One of the final steps in annotating a signal is to verify that the annotation labels match your normal conventions for labeling. For example, you may want to check that only labels from a given inventory are present. Another step in the final processing may be to collapse adjacent silences/pauses into single labels.

These kinds of operation can be most easily performed with an SML script. We will present two scripts: the first checks labels against an inventory stored in a file, the second collapses silences and pauses.

Verification

We assume that an inventory of symbols is saved in a text file with one symbol per line. The following script then reports the name and location of all symbols not in the inventory.

```
/* anverify - verify annotation labels come from known inventory */

/* inventory */
file    ip;
string    itab[1:1000];
var    icnt;

/* load inventory from file */
init {
   string    s;
   openin(ip,"c:/sfs/dev/sampa.lst");

   input#ip s;
   while (compare(s,s)) {
      icnt = icnt+1;
      itab[icnt] = s;
      input#ip s;
   }

   close(ip);
}

/* process an annotation item */
main {
   var    i,num;

   num = numberof(".");
   for (i=1;i<=num;i=i+1) {
      if (!entry(matchn(".",i),itab)) {
                                    print $filename,"\t";
         print timen(".",i):8:4,"\t";
         print matchn(".",i)," – illegal symbol\n";
      }
   }
}
```

To run this script, copy and paste it into a file "anverify.sml" and create the inventory file "sampa.lst". Then select the annotation item to check and run Tools|Run SML script, see Figure C.5.1.

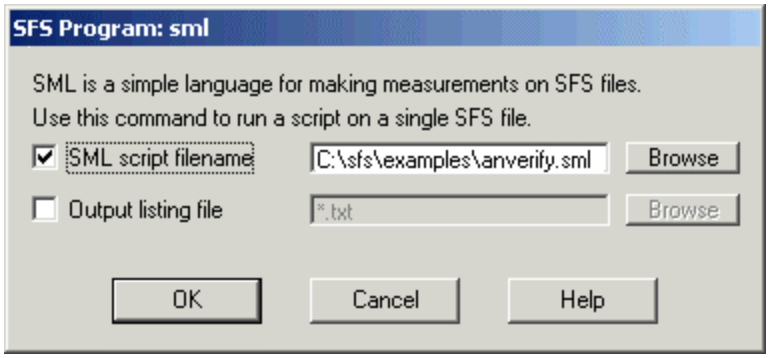


Figure C.5.1 - SFSWin Run SML script dialog

Post-processing

In this script we collapse adjacent annotations if they both label silence or pause. Specifically:

First	Second	Result
...	...	...
...	/	...
/	...	...
/	/	/

The processed annotation item is saved back into the same file.

```
/* ansilproc - collapse adjacent silence annotations */

item    ian;    /* input annotations */
item    oan;    /* output annotations */

/* check annotation for silence */
function var issil(lab)
string lab
{
    if (compare(lab,"/")==0) return(1);
    if (compare(lab,"...")==0) return(1);
    return(ERROR);
}

main {
    var    i,j,numf;
    var    size,cnt;
    string    lab,lab2;

    /* get input & output */
    sfsgetitem(ian,$filename,str(selectitem(AN),4,2));
    numf=sfsgetparam(ian,"numframes");
    sfsnewitem(oan,AN,sfsgetparam(ian,"frameduration"),
        sfsgetparam(ian,"offset"),1,numf);

    /* process annotations */
    i=0;
    cnt=0;
    while (i < numf) {
        lab = sfsgetstring(ian,i);
        if ((i<numf-1) && issil(lab)) {
            /* is a non-final silence */
            size=sfsgetfield(ian,i,1);
            j=i+1;
```

```
            lab2 = sfsgetstring(ian,j);
            while ((j<numf) && issil(lab2)) {
                if (compare(lab2,"...")==0) lab = lab2;
                size=size + sfsgetfield(ian,j,1);
                j=j+1;
                if (j<numf) lab2 = sfsgetstring(ian,j);
            }
            sfssetfield(oan,cnt,0,sfsgetfield(ian,i,0));
            sfssetfield(oan,cnt,1,size);
            sfssetstring(oan,cnt,lab);
            i=j;
        }
        else {
            /* final or non-silence, just copy */
            sfssetfield(oan,cnt,0,sfsgetfield(ian,i,0));
            sfssetfield(oan,cnt,1,sfsgetfield(ian,i,1));
            sfssetstring(oan,cnt,lab);
            i=i+1;
        }
        cnt = cnt + 1;
    }

    /* save result */
    sfsputitem(oan,$filename,cnt);
}
```

Copy and paste this script into ansilproc.sml, and run it using Tools|Run SML script. An example of the effect of the script is shown in Figure C.5.3

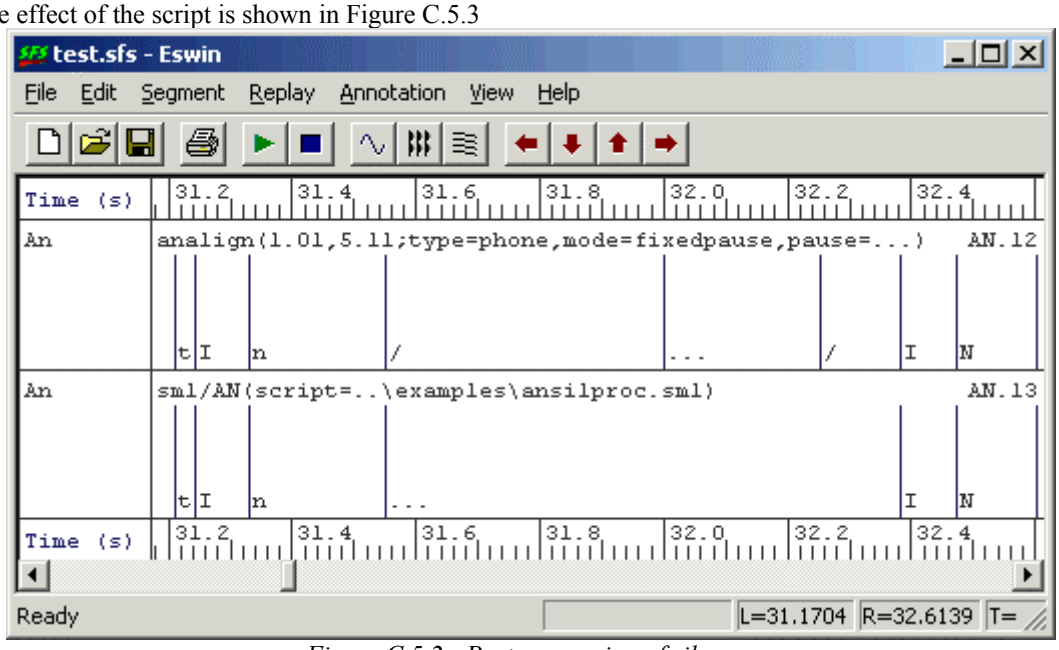


Figure C.5.2 - Post-processing of silences

6. Other special processing

This section refers specifically to annotated recordings of dysfluent speech made available by the Speech Group of the Department of Psychology at UCL (www.psychol.ucl.ac.uk).

Description of UCL Psychology phonetic annotation system

Below is a summary of the phonetic mark-up developed by the Speech group and used on the dysfluent speech database. The basic phonetic symbol set is the JSRU symbol set described in Appendix B.

Word Boundaries

Word boundaries are indicated in the phonetic transcription with a symbol placed before the first syllable in the word:

- A forward slash "/" is used to mark a function Word
- A colon ":" is used to mark a content Word

Function words are closed class words (only about 300 in English) which perform grammatical functions while content words are open class words which carry meaning.

Function Words	
Prepositions:	of, at, in, without, between
Pronouns:	he, they, anybody, it, one
Determiners:	the, a, that, my, more, much, either, neither
Conjunctions:	and, that, when, while, although, or
Modal verbs:	can, must, will, should, ought, need, used
Auxilliary verbs:	be (is, am, are), have, got, do
Particles:	no, not, nor, as
Content Words	
Nouns:	John, room, answer, Selby
Adjectives:	happy, new, large, grey
Full verbs:	search, grow, hold, have
Adverbs:	really, completely, very, also, enough
Numerals:	one, thousand, first
Interjections:	eh, ugh, phew, well
Yes/No answers:	yes, no (as answers)

Beware that the same lexical word can function as either content or function word depending on its function in an utterance:

1. have
    a. "I have come to see you" = Function Word (Auxillary)
    b. "I have three apples" = Content Word (Full Verb)
2. one
    a. "One has one's principles" = Function Word (Pronoun)
    b. "I have one apple" = Content Word (Numeral)
3. no
    a. "I have no more money" = Function Word (Negative Particle)
    b. "No. I am not coming" = Content Word (Yes/No Answer)

Examples with the word boundary markers:

- "I saw him in the school." = /ie :saw /him /in /dha :skuul.
- "I have come to see you." = /ie /haav :kam /ta :see /yuu.

Syllable Boundaries

The appropriate stress marker from the list below is placed at the start of each syllable, to mark syllable boundaries as well as stress:

- Exclamation mark ! prior to emphatically-stressed syllable.
- Double quote " prior to primary-stressed syllable.
- Single quote ' prior to secondary-stressed syllable.

- Hyphen - prior to unstressed syllable not in word-initial position.

In the case of a word-initial syllable, the stress marker is positioned immediately after the word marker. The only exception is in the case of an unstressed first syllable, which does not receive a dash but instead only receives the word marker. The dash, by default, indicates that a syllable is not word initial, as well as indicating that it is unstressed.

Examples:

- :"sen-ta (centre)
- :di"tekt (detect)
- :'in-ta"naa-sha-nl (international)
- :"in-ta'naa-sha-na-lie-zai-shn (internationalization)
- :in-ta'naa-sha-na-lie"zai-shn (internationalization)

Dysfluencies within words

All dysfluent phones are entered in UPPER-CASE at a finer-grained level of transcription wherein each upper-case symbol represents 50ms duration estimates.

Multiple upper-case phones may be represented with an explicit repetition count: {x *num*}, e.g. if the duration of a prolonged F were 5 times 50ms, it could be transcribed either as "FFFFF" or F{x 5}. The latter is helpful in transcribing very long prolongations like F{x 30}.

A "Q" is used to indicate a pause within a word of 100ms, e.g. a 300ms dysfluent pause would be transcribed as either QQQ or Q{x 3}.

Examples:

- Prolongations, e.g. /dhaats :FFFFFaan"taa-stik or /dhaats :F{x 5}aan"taa-stik.
- Repetitions, e.g. dhaats /a :"load /av :"BA BA BA BA Bawl-da'daash or /dhaats /a :"load /av :"KQ KQ KQ Ko-bl-az.

Note that a space does not indicate pausing. In the first repetition example above, there is no pause between the repetitions of the "BA" sound. There are, however, brief pauses (100ms) between the "K" sounds in the second example

For ambiguous phonetic transcription sequences the {x…} convention is used when the symbol is repeated, e.g. the transcription "AAAAA" refers to a prolonged "A", but "AA{x5}" refers to a prolonged "AA".

Other dysfluencies which cannot be transcribed are entered in the form of a comment at the place where it occurs. For example, a block can be entered as {U block}. All dysfluencies are marked in the phonetic transcriptions.

Marking of supralexical dysfluencies

Word repetitions are transcribed using the syllable or word repetition convention described below (++|++), with the exception that a monosyllabic word that is repeated with no pausing, or very little, and is judged to be 'stuttered' can be transcribed within one word, thus:

- /AAND /aand, or, /AANDQQ/aand

Any repeated monosyllabic words that are separated by significant pausing (more than two Qs) are transcribed using the convention below.

In the transcribed speech, the section of "replaced" and "replacement" speech are enclosed by two "+" signs and the two sections are separated by a vertical bar "|". For example:

- Syllable repetition:

  /dhei /waz :"noa + :"ree Q + | + :"ree + -zan /fa /him ta :"duu /it.

- Word repetition:

  /dheiz :"noa :"u-dha + /dhat + | + /dhat + /ie :'noa /ov.

- Backtracking:

  /dheiz + :"noa :"u-dha /dhat Q + | + :"noa :"u-dha /dhat + /ie :'noa/ov.

- Backtracking + elaboration:

/dheiz + :"noa :"u-dha /dhat Q + | + :"noa :"u-dha :'i-di-at /dhat + /ie :'noa /ov.

Marking of pauses between words

Pauses are marked with lower-case "q" if they are part of fluent speech intonation. Dysfluent pauses are marked with upper-case "Q".

Marking of other comments

Other comments by the transcriber are entered into the transcription using the convention {U ...text...}. This might be used for speech that could not be transcribed or for other sound events.

Division into tiers

In this section we will look at how the dysfluent transcription may be divided into two tiers: the first describing the word and dysfluency events, the second describing the phonetic sequence. The advantage of this separation is that the phonetic symbol sequence may then be time-aligned with the speech signal. Figure C.6.1 shows an example of the database annotation prior to processing.

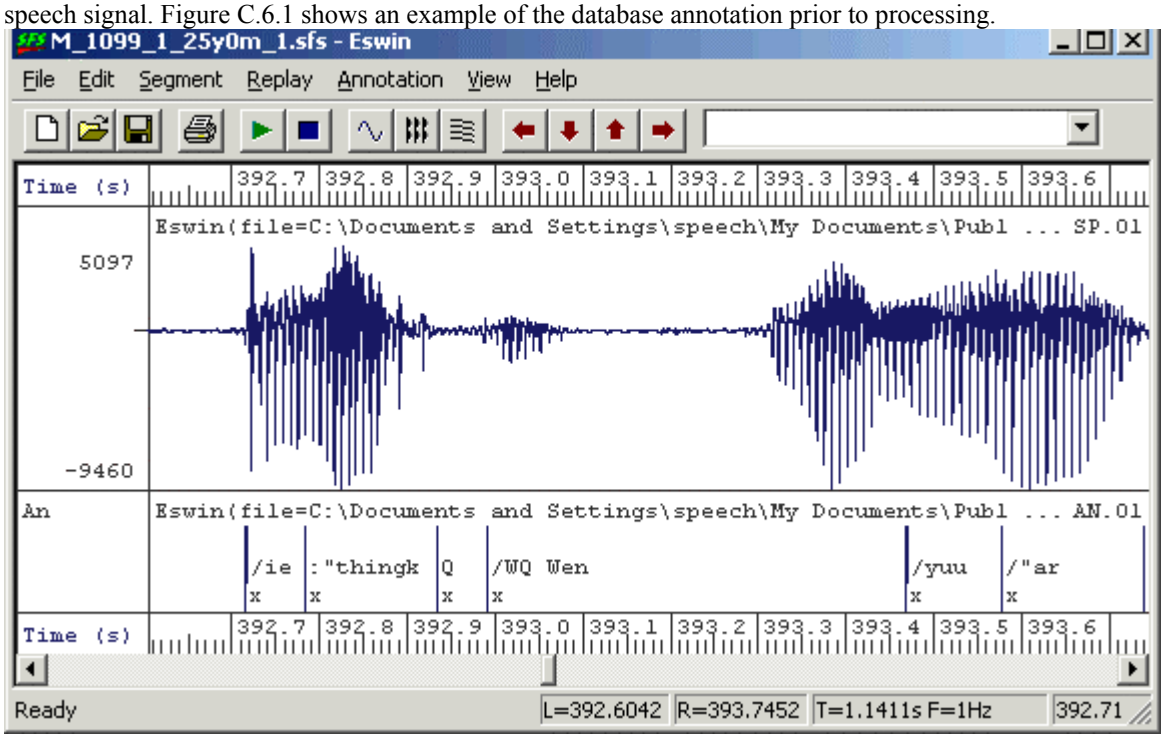


Figure C.6.1 - Example dysfluency mark-up before processing

The following script takes as input an annotation item marked up using the system above which has been time-aligned at the level of individual syllables, as can be seen in Figure C.6.1. The output of the script is two further annotation sets. The basic principle of operation is that the input transcription is parsed symbol by symbol, while some symbols are directed into the word tier and others into the phone tier. In addition, redundant annotations that only mark the ends of syllables (and are shorter than about 5ms long) are removed.

The marking of dysfluency is changed so that conventional phonetic annotation is used in the phone tier, and a new marker "{D}" is added to the word tier. This allows us to process the phone tier using the automatic alignment procedure described in section 4.

At present no processing of "multiplier" markers is performed, so that "AA{x 5}" is divided up into "aa" on the phone tier and "{x 5} {D}" on the word tier.

```
/* anfluency - process phonetic annotations used on fluency data */

/* version 1.0 - June 2004
 *
 * This script take a set of phonetic annotations
 * from the UCL Psychology Speech Group fluency
 * database and normalises them to be consistent
 * with SFS conventions.
 *
```

```
* Input is transcription of syllables or words
* in JSRU format with additional markers showing
* word category and dysfluency
*
* Output is two new annotation sets: one containing
* only the phonetic labels and stress markers, parsed
* with spaces between the symbols; and one with the
* word category and dysfluency mark-up
*
*/


item    oanp;    /* output annotations - phonetic */
var     oanpcnt;
item    oanw;    /* output annotations - word */
var     oanwcnt;

/* check for uppercase */
function var isupper(str)
{
   string str;
   if ((ascii(str)>=65)&&(ascii(str)<=90)) return(1);
   return(ERROR);
}

/* convert to lower case */
function string tolower(src)
{
   string src;
   string dst;
   var       i;
   dst="";
   for (i=1;i<=strlen(src);i=i+1) {
      if ((ascii(src:i:i)>=65)&&(ascii(src:i:i)<=90)) {
         dst = dst ++ char(ascii(src:i:i)+32);
      }
      else {
         dst = dst ++ src:i:i;
      }
   }
   return(dst);
}

/* check next character for digraph */
function string checknext(prefix,ch,label)
string label;
{
   string prefix,ch;
   if (strlen(label)==0) return(prefix);
   if (index(ch,label:1)==1) {
      prefix=prefix++(label:1);
      label=label:2:strlen(label);
   }
   return(prefix);
}

/* strip next symbol from front of string */
function string nextsymbol(label)
string label;
{
```

```
string c;
string l;
var idx;

while (1) {
   /* strip off first character */
   if (strlen(label)==0) return("");
   if (strlen(label)==1) {
      c=label;
      label="";
   }
   else {
      c=label:1;
      label=label:2:strlen(label);
   }

   /* action based on character */
   switch (c) {
   case " ":    { /* skip */ }
   case "_":    { /* skip */ }
   case "~":    { /* skip */ }
   case "/":    return(c);
   case ":":    return(c);
   case "(":    {  /* dysfluency mark-up */
                /* these should probably be mapped to {} */
                idx=index("\)",label);
                if (idx) {
                   l=c++(label:idx);
                   label=label:idx+1:strlen(label);
                   return(l);
                }
                else {
                   l=c++label;
                   label="";
                   return(l);
                }
             }
   case "{":    {  /* dysfluency mark-up */
                idx=index("}",label);
                if (idx) {
                   l=c++(label:idx);
                   label=label:idx+1:strlen(label);
                   return(l);
                }
                else {
                   l=c++label;
                   label="";
                   return(l);
                }
             }
   pattern "[aA]":    return(checknext(c,"[airwAIRW]",label));
   pattern "[cC]":    return(checknext(c,"[hH]",label));
   pattern "[dD]":    return(checknext(c,"[hH]",label));
   pattern "[eE]":    return(checknext(c,"[eiryEIRY]",label));
   pattern "[gG]":    return(checknext(c,"[xX]",label));
   pattern "[iI]":    return(checknext(c,"[aeAE]",label));
   pattern "[nN]":    { /* special processing for sequence "ngx" => "n gx" */
                      l = checknext(c,"[gG]",label);
                      if ((compare(l,"ng")==0)&&(compare(label:1,"x")==0)) {
                         /* mis-parse ngx */
```

```
                        label="g"++label;
                        return("n");
                    }
                    return(l);
                }
        pattern "[oO]":    return(checknext(c,"[aiouAIOU]",label));
        pattern "[sS]":    return(checknext(c,"[hH]",label));
        pattern "[tT]":    return(checknext(c,"[hH]",label));
        pattern "[uU]":    return(checknext(c,"[ruRU]",label));
        pattern "[zZ]":    return(checknext(c,"[hH]",label));
        default:    return(c);
        }
    }
}

/* process a chunk of dysfluent transcription */
function var processlabel(posn,size,label)
{
    var    posn;
    var    size;
    string label;
    string    sym;
    string    plabel;
    string    wlabel;
    var       idx;
    var       dysfluent;

    /* check valid label */
    if (!compare(label,label)) return(0);

    /* initialise */
    sym=nextsymbol(label);
    plabel="";
    wlabel="";
    dysfluent=0;

    /* while symbols left */
    while (compare(sym,"")!=0) {
        if (index("[:/\({]",sym:1)) {
            /* is word type or dysfluency mark-up */
            wlabel=wlabel++" "++sym;
        }
        else if (index("[Q]",sym:1)) {
            /* is pause - add to both tiers */
            wlabel=wlabel++" "++sym;
            plabel=plabel++" q";
        }
        else {
            if (isupper(sym)) {
                /* is dysfluent phone */
                plabel=plabel++" "++tolower(sym);
                dysfluent=1;
            }
            else {
                /* is normal phone */
                plabel=plabel++" "++sym;
            }
        }
        sym=nextsymbol(label);
    }
```

```
   /* add dysfluent marker to word tier */
   if (dysfluent!=0) wlabel = wlabel ++ " " ++ "{D}";

   if (strlen(wlabel)>1) {
      /* add word tier label */
      sfssetfield(oanw,oanwcnt,0,posn);
      sfssetfield(oanw,oanwcnt,1,size);
      sfssetstring(oanw,oanwcnt,wlabel:2:strlen(wlabel));
      oanwcnt=oanwcnt+1;
   }
   /* add something to phone tier */
   if (strlen(plabel)==0) plabel=" q";
   sfssetfield(oanp,oanpcnt,0,posn);
   sfssetfield(oanp,oanpcnt,1,size);
   sfssetstring(oanp,oanpcnt,plabel:2:strlen(plabel));
   oanpcnt=oanpcnt+1;
}

/* for each input file */
main {
   string   anitemno;
   var      i,j,numf;
   var      posn,posn2,size;
   string   lab,lab2;
   var      eps,pause;

   /* get input annotation set and made output annotation sets */
   anitemno=str(selectitem(AN),4,2);
   sfsgetitem(ian,$filename,anitemno);
   numf=sfsgetparam(ian,"numframes");
   sfsnewitem(oanp,AN,sfsgetparam(ian,"frameduration"),\
      sfsgetparam(ian,"offset"),1,numf);
   sfssetparamstring(oanp,"history",\
      "sml("++anitemno++";script=anfluency.sml,type=phone)");
   sfsnewitem(oanw,AN,sfsgetparam(ian,"frameduration"),\
      sfsgetparam(ian,"offset"),1,numf);
   sfssetparamstring(oanw,"history",\
      "sml("++anitemno++";script=anfluency.sml,type=word)");

   /* processing constants: small time (6ms) and pause time (100ms) */
   eps = trunc(0.5 + 0.006 / sfsgetparam(ian,"frameduration"));
   pause = trunc(0.5 + 0.1 / sfsgetparam(ian,"frameduration"));

   /* process annotations */
   oanwcnt=0;
   oanpcnt=0;
   for (i=0;i<numf;i=i+1) {
      /* get input annotation */
      lab = sfsgetstring(ian,i);
      posn = sfsgetfield(ian,i,0);
      size = sfsgetfield(ian,i,1);

      /* check is not a redundant 'x' */
      if ((i<numf-1)&&(compare(lab,"x")==0)) {
         posn2 = sfsgetfield(ian,i+1,0);
         if (posn2 > posn+eps) {
            /* next annotation far off - insert / */
            sfssetfield(oanp,oanpcnt,0,posn);
            sfssetfield(oanp,oanpcnt,1,size);
```

```
            if (size < pause) lab="q" else lab="...";
            sfssetstring(oanp,oanpcnt,lab);
            oanpcnt = oanpcnt+1;
            /* also copy long pauses into word tier / */
            if (size >= pause) {
               sfssetfield(oanw,oanwcnt,0,posn);
               sfssetfield(oanw,oanwcnt,1,size);
               sfssetstring(oanw,oanwcnt,"...");
               oanwcnt = oanwcnt+1;
            }
         }
      }
      else if (compare(lab:1,"x")==0) {
         /* annotation starting with x - strip x */
         if (strlen(lab)>1) processlabel(posn,size,lab:2:strlen(lab));
      }
      else {
         /* normal annotation */
         processlabel(posn,size,lab);
      }
   }

   /* report processing */
   print $filename,": processed ",numf:1," annotations into "
   print oanwcnt:1," word and ",oanpcnt:1," phone annotations\n";

   /* save results */
   if (oanwcnt > 0) sfsputitem(oanw,$filename,oanwcnt);
   if (oanpcnt > 0) sfsputitem(oanp,$filename,oanpcnt);

}
```

An example of the processing performed by the script can be seen in Figure C.6.2.

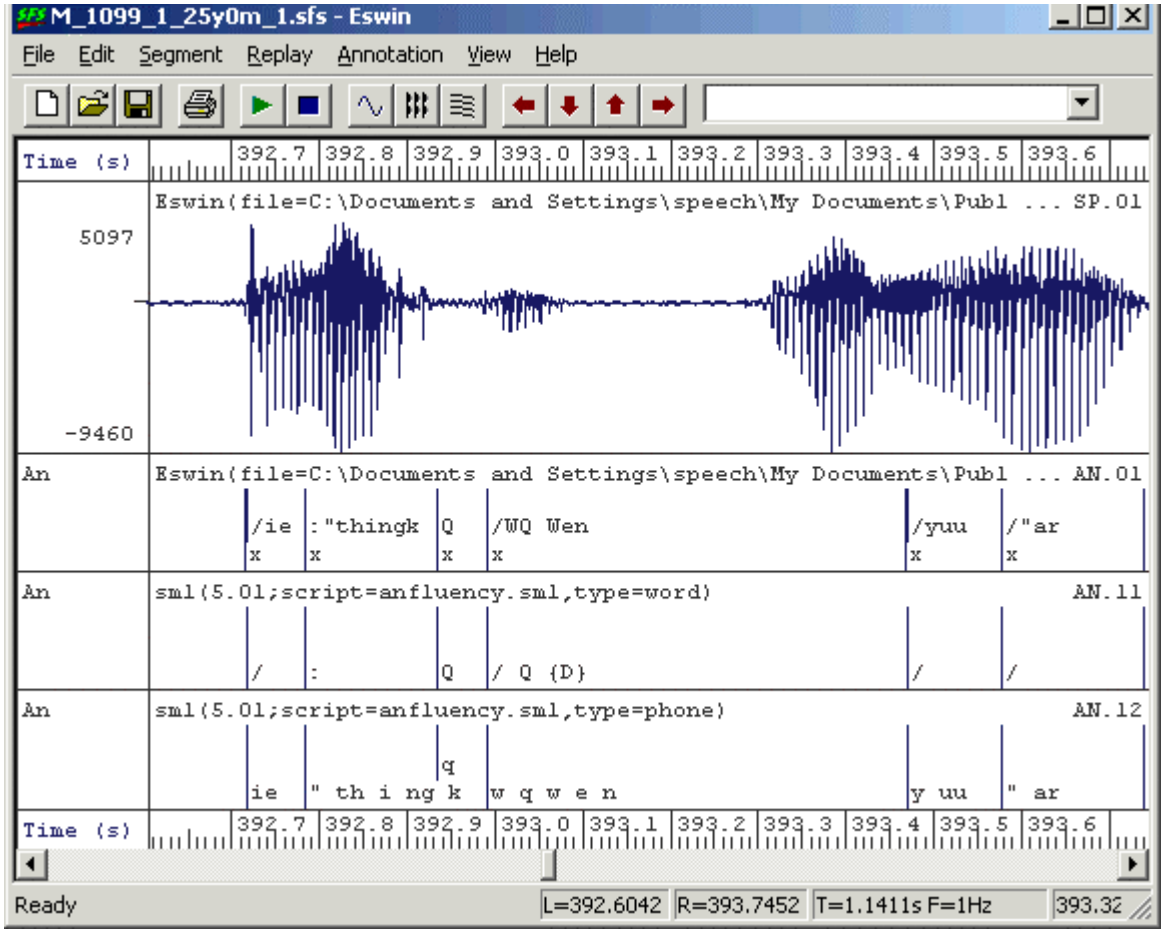


Figure C.6.2 - Example dysfluency mark-up divided across two tiers

Subsequent Processing

Phonetic alignment of phone tier

The automatic phonetic alignment of the phone tier can be performed using the tools describe in section 4 of this appendix. Selecting a suitable speech and annotation item, choose menu option Tools|Annotations|Auto align phone labels. For the phone tier annotations above we only want to align within a phone label and to use JSRU symbols. See Figure C.6.3.

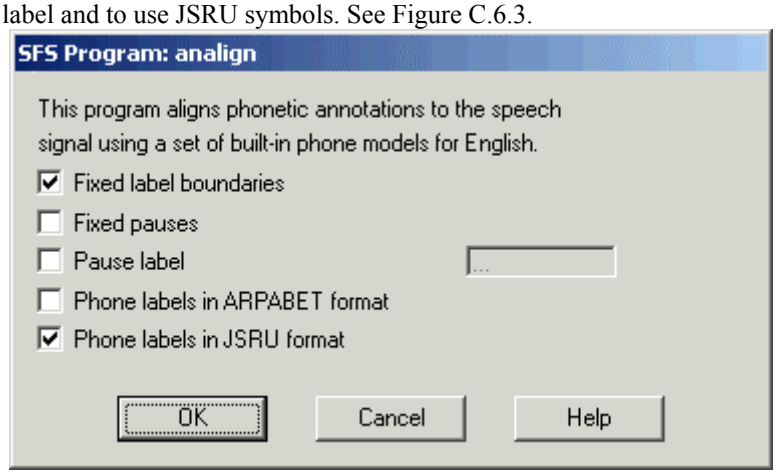


Figure C.6.3 - Automatic phonetic alignment on JSRU symbols

Dysfluency statistics

Finally, we will show how the identification of dysfluencies in the word tier can be used to collect some statistics about their occurrence. This script counts where dysfluencies occurred and their typical duration.

```
/* dysstats.sml - measure some statistics about dysfluencies */
```

```
/* counts */
var    ncontent;    /* # in content words */
var    nfunction;    /* # in function words */
var    npause;       /* # after pause */

/* stats */
stat    sdur;       /* stats on duration */

/* for each input file */
main {
   var    num,i;
   string    last;
   string    lab;

   num=numberof(".");

   last="";
   for (i=1;i<=num;i=i+1) {
      lab=matchn(".",i);
      if (index("{D}",lab)) {
         if (index("/",lab)) {
            nfunction=nfunction+1;
         }
         else if (index(":",lab)) {
            ncontent=ncontent+1
         }
         if (index("Q",last)) {
            npause=npause+1;
         }
         sdur += lengthn(".",i);
      }
      last=lab;
   }
}

/* summarise */
summary {
   print "Files processed          : ",$filecount:1,"\n";
   print "Number of dysfluencies   : ",nfunction+ncontent:1,"\n";
   print "Dysfluent function words : ",nfunction:1,"\n";
   print "Dysfluent content words  : ",ncontent:1,"\n";
   print "Dysfluencies after pause : ",npause:1,"\n";
   print "Mean dysfluent duration  : ",sdur.mean," +/- ",sdur.stddev,"s\n";
}
```

An example run of the script on one 15 min recording is shown below:

```
Files processed          : 1
Number of dysfluencies   : 28
Dysfluent function words : 18
Dysfluent content words  : 10
Dysfluencies after pause : 7
```

Bibliography

- BEEP British English pronunciation dictionary at ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/beep.tar.gz.
- Hidden Markov modeling toolkit at http://htk.eng.cam.ac.uk/.

- International Phonetic Alphabet at http://www.arts.gla.ac.uk/IPA/ipachart.html.
- SAMPA Phonetic Alphabet http://www.phon.ucl.ac.uk/home/sampa/.
- Speech Filing System at http://www.phon.ucl.ac.uk/resource/sfs/.
- UCL Psychology Department at http://www.psychol.ucl.ac.uk/.
- VoiScript at http://www.phon.ucl.ac.uk/resource/voiscript/.

Appendix D Audio analysis with SFS

Researchers might be interested in the way that production of particular sounds is affected by a disorder such as stammering. They may resort to audio analysis to do this. This Appendix describes some basic ways in which audio analysis of speech can be performed using SFS utilities. The main topic covered is formant analysis, which is a way of representing how speech output changes over time as the articulators move to produce the consonant and vowel sounds described in Appendix B. Some background on 1) articulatory phonetics and spectrographic analysis of speech and 2) statistical analysis are assumed. For readers needing the background for, or those who wish to revise the concepts behind 1,) reference can be made to Rosen and Howell (1991) which provides an elementary, non-mathematical introduction to this area. Hyperlinks to websites that cover some of the critical concepts behind the statistical topics are given at the end. Some software is presented and described in this appendix but programming experience is not assumed. This tutorial refers to versions 4.6 and later of SFS and appears in the documentation on the SFS website. Visit the SFS website to obtain your software (http://www.phon.ucl.ac.uk/)

1. Formant Analysis Strategy

Perhaps the most obvious way to do formant analysis with SFS is to load up an audio signal, choose Tools|Speech|Display|Cross-section, then make measurements of formant frequencies interactively, writing the results down on a piece of paper, see figure D.1.1. You can then type your results into a statistics package and make whatever comparisons you need. This is **not** the strategy we will be presenting in this tutorial.

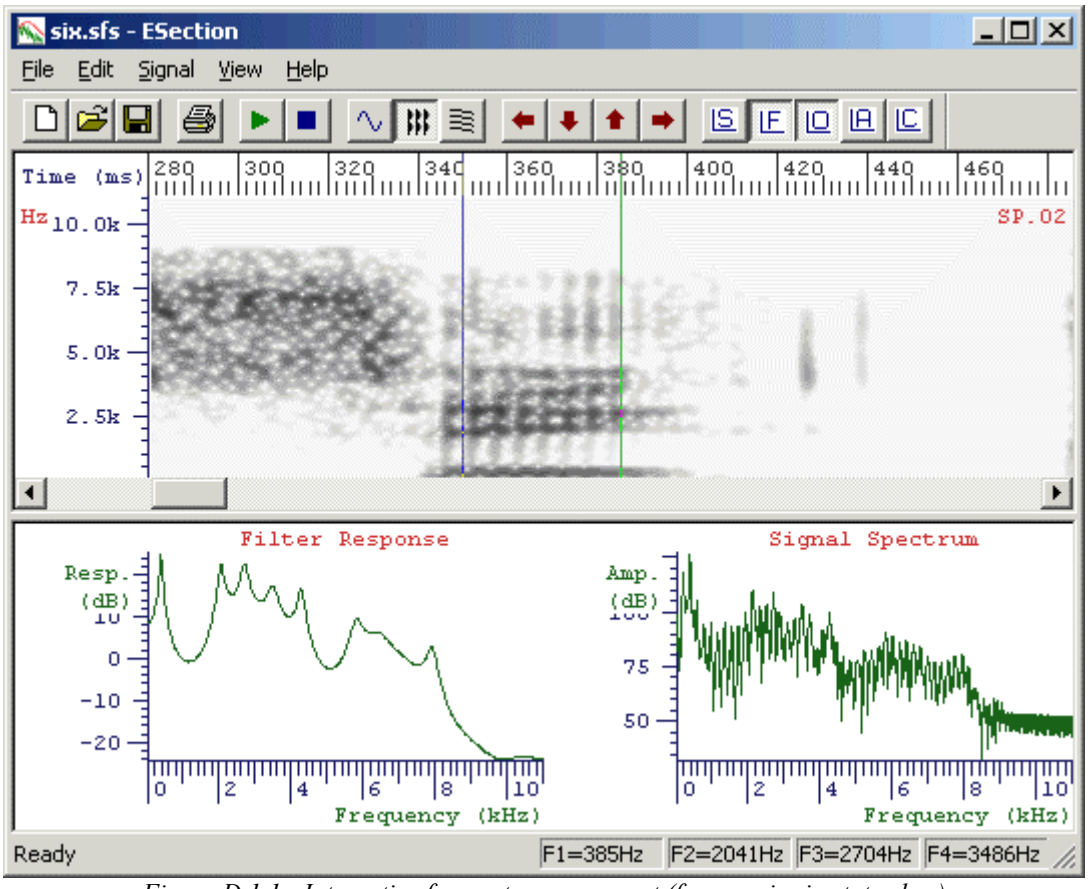


Figure D.1.1 - Interactive formant measurement (frequencies in status bar)

There are a number of deficiencies in the direct, interactive route:

- Lack of consistency: how do you know that you are positioning the cursors in a consistent fashion every time?
- Potential bias: are you sure you haven't chosen the position of the cursors to obtain the expected formant frequency values?

- Inflexibility: if you want to go back and change the way the formants are measured, or if you want to go back and collect other data (e.g. durations), you have to go through all the data again *without knowing exactly how you made the first set of measurements*.

- Cost: measuring interactively is slow and time-consuming. If you can use a speech corpus that has been phonetically labeled, then you can be much more productive by exploiting those labels.

- Amount of data: a semi-automatic procedure can analyse more data and provide a wider range of statistics on the distribution of formant frequencies.

The strategy we will be presenting in this tutorial follows these steps:

1. Annotate the signal;

2. Perform automatic formant analysis of all the speech data;

3. Use a script to extract the distribution of formant values;

4. Analyse the formant distributions.

2. Annotating the audio signal

If you are lucky you will find ready-annotated material suitable for your purposes (11 such files from speakers who stammer are available from the UCL data archive described in Appendix A). Phonetically-annotated speech corpora are becoming more common, though they are still rare for speech disorders like stammering. You will probably also require data on fluent speakers for control purposes and if you are thinking of analysing one of the major languages of the world you should investigate whether annotated recordings are available for fluent speakers at least. Often these are supplied to speech researchers at a much lower price than they are made available to speech technology companies. The two major world suppliers of speech corpora are the Linguistic Data Consortium (www.ldc.upenn.edu) and the European Language Resource Association (www.elra.info).

We will not concern ourselves here with converting corpus data to be compatible with SFS, but there exist tools in SFS (such as cnv2sfs and anload) to help make this easier. Instead we will briefly discuss the use of SFS to add annotations to the signal. We assume that we will only be annotating the regions of the signal where we want to make formant measurements, rather than performing an aligned phonetic annotation of the whole signal (see the sister tutorial in Appendix C on Phonetic annotation).

Typically formant measurements are made on syllabic nuclei, where there is likely to be voicing and a relatively unobstructed vocal tract. We will describe the annotation of monophthongal vowels, although the procedure could easily be adapted to deal with more complex elements.

From within SFSWin, select the speech item to annotate and choose Tools|Speech|Annotate|Manually. Enter a suitable name for the annotations (say, "labels"). The speech signal will be displayed, with a box at the bottom of the display where the annotations may be added/edited. Adjust the display to show waveforms and/or spectrograms as you wish. Use the cursors to isolate regions of the signal until you find the first vowel segment you want to annotate. Zoom in so that the vowel region is clearly visible - perhaps filling about one-quarter of the display. See Figure D.2.1.

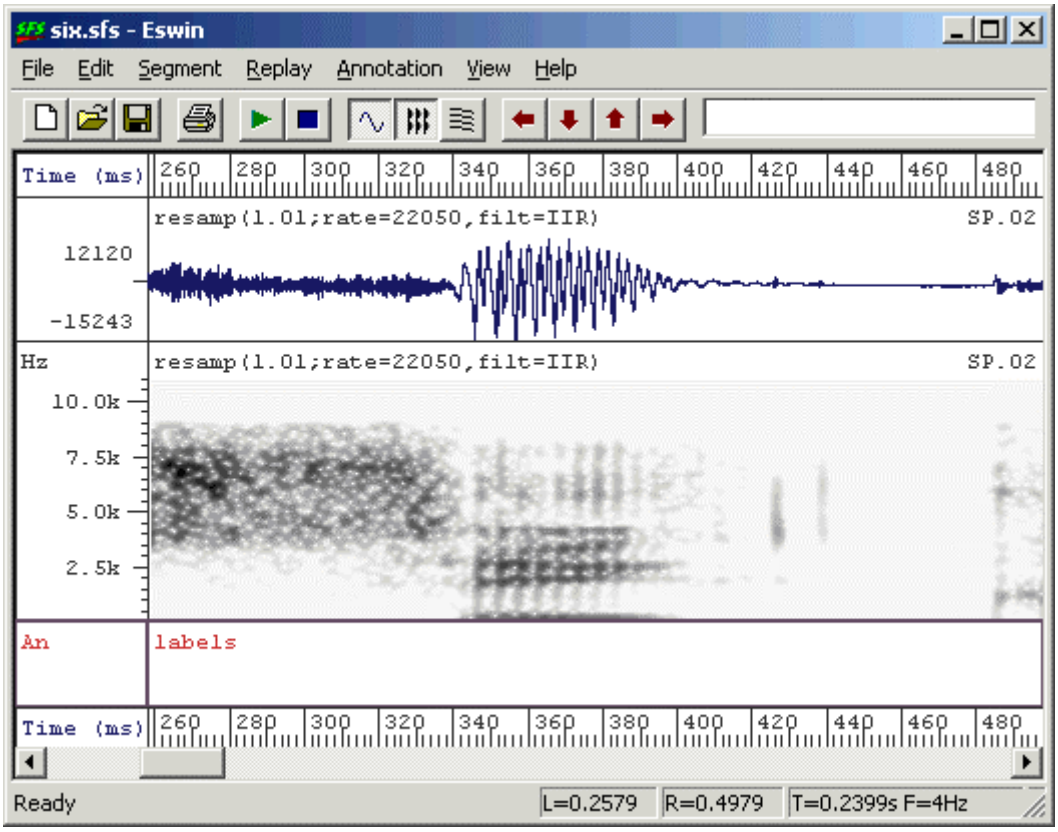


Figure D.2.1 - Ready for annotation

The next question is how best to position the annotations. Should one attempt to determine the "centre" of the vowel, or its "edges", or where the formants are "stationary"? The problem is that none of these have clear, unambiguous definitions. The best choice is the one that makes the least assumptions and has the least potential for bias. It is suggested that a strategy for labeling is chosen that is easy and reliable. In the case of Figure D.2.1, where the vowel is preceded and followed by a voiceless consonant, then the labels should go at the start and end of voicing. In circumstances where the voiced region is shared with another voiced segment, consider estimating the point which is acoustically half-way between the segments. It is then reasonable to propose that one segment dominates on one side of the label, while the other segment dominates on the other side. Once the labels are positioned we can try various programmed strategies for reliably extracting formant frequencies from the region.

To add an annotation, position the left cursor at the start of the region to annotate, and the right cursor at the end. Then using the keyboard, type the following:

[A] [*label*] [RETURN]
[B] [/] [RETURN]

The [A] key is a keyboard shortcut meaning label the left cursor, the [B] key is a keyboard shortcut meaning label the right cursor. In Figure D.2.2 we have labeled a segment as being "I_six", that is the vowel /I/ in the word "six". We have marked the end of the segment with "/".

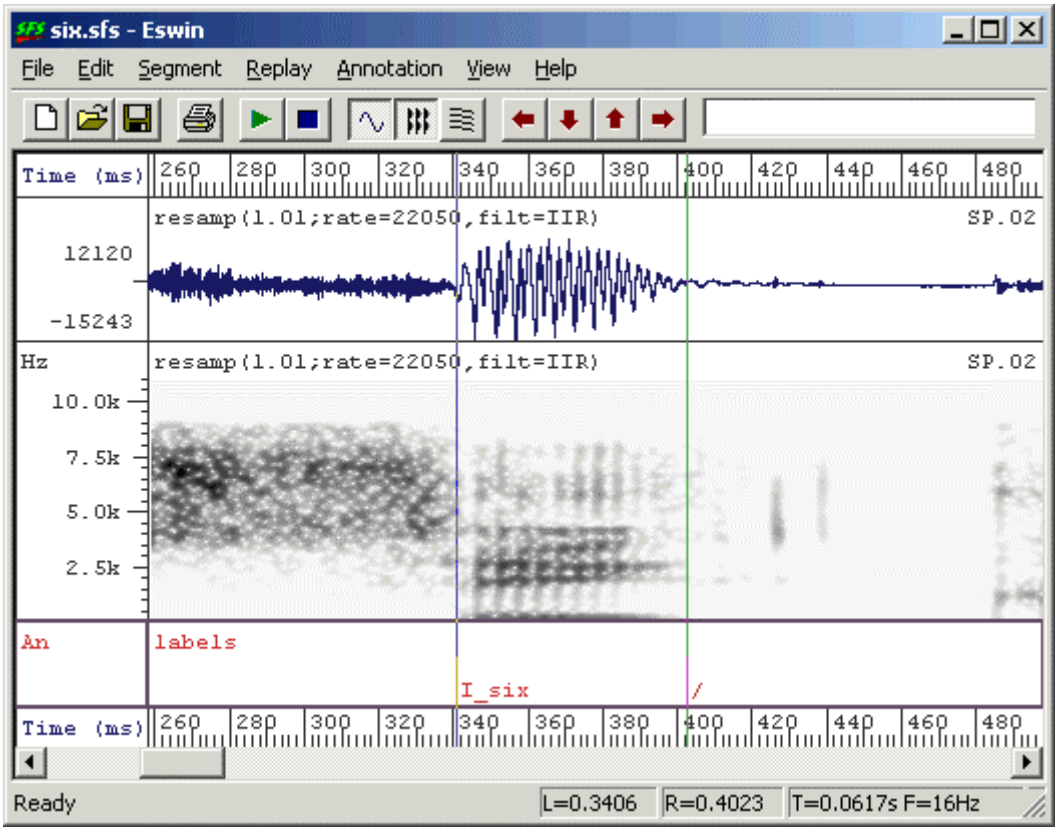


Figure D.2.2 - Annotated vowel segment

Deciding how to label the regions will depend on the kind of phonetic analysis you are planning to undertake. Since adding information to the label is easy when you are doing the labeling, and very difficult to do retrospectively, consider labeling with information about the context as well as the identity of the segment. You may find in your analysis that your assumptions about allophonic variation are wrong, and that identically-labelled segments actually belong to different phonological categories!

One final useful piece of advice is to fill in information about the identity of the speaker and the recording session in the SFS file header. This information may be useful in allowing us to find files and label graphs later. To do this, select option File|Properties in SFSWin and complete the form shown in Figure D.2.3.

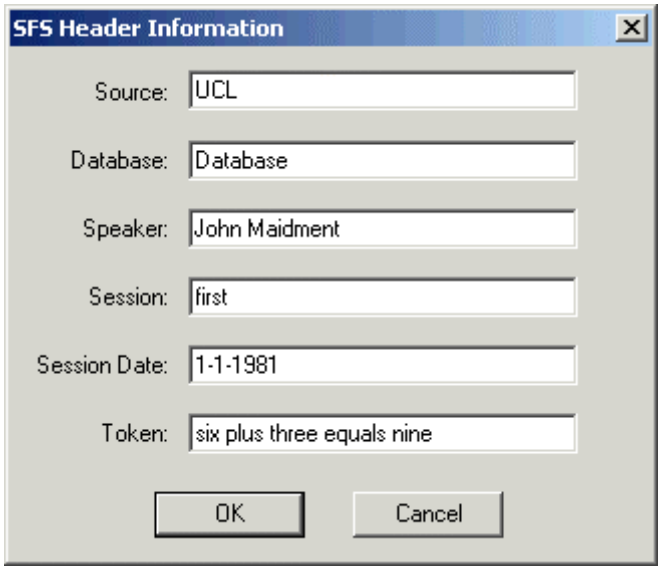


Figure D.2.3 - SFSWin File properties

3. Formant Analysis

Introduction

It is worth mentioning at the outset that formant analysis is not an exact science. The task the computer is trying to do is to estimate the natural frequency of vocal tract resonances given a short section of speech signal picked up by a microphone. The task is made complex because the only way information about the resonances gets into the microphone signal is if the resonances are excited with sounds generated elsewhere in the vocal tract - typically from the larynx. Thus the program has to make assumptions about the nature of this source signal to determine how that signal has been modified by the vocal tract resonances. It may be the case that peaks in the spectrum of the sound are caused by vocal tract resonances, but they may be properties of the source. Likewise, it may be the case that every formant is excited, but it may be that the source simply had no energy at a resonant frequency and it was not excited.

In addition formant analysis is made difficult by the following factors: the articulators are constantly moving and the source is changing while producing speech; the sound signal generated by the vocal tract is possibly contaminated by noise and reverberation before it enters the microphone: and sometimes formants can get very close together in frequency - so that two resonances can give rise to a single spectral peak. All this without even mentioning the problems that arise at high fundamental frequencies, when formant frequencies are likely to be biased towards the nearest harmonic frequency.

In all, we should expect that our formant analysis will give rise to "errors", and rather than ignoring them, or hoping that they have no effect, we should build in the possibility of measurement error into our procedures.

Fixed frame analysis

The most common means to obtain formant frequency measurements from a speech signal is through linear prediction on short fixed-length sections of the signal - typically 20-30ms windows. These windows are stepped by 10ms to give spectral peak estimates 100 times per second of signal. Typically each frame delivers about 6 spectral peaks from a signal sampled at 10,000 samples/sec. Not all these peaks are caused by formants, and so a post-processing stage is required to label some of the peaks as being caused by "F1", "F2", etc. This post-processing stage usually makes assumptions about the typical frequency and bandwidth range of vocal tract resonances and their rate of change.

Currently the best program in SFS to perform fixed-frame formant analysis is the formanal program. This can be found in SFSWin under Tools|Speech|Analysis|Formant estimates track. The formant analysis code in this program was originally written by David Talkin and John Shore as part of the Entropic Signal Processing System and is used under licence from Microsoft. The current SFS implementation does not have any user-changeable signal processing parameters, see Figure D.3.1.

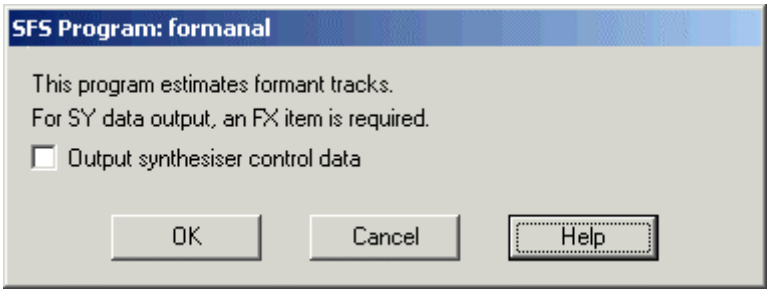


Figure D.3.1 - SFSWin Formant estimates dialog

The formanal program performs the following processing steps:

1. Downsamples the signal to 10,000 samples/sec;
2. High-pass filters at 75Hz;
3. Pre-emphasises the signal;
4. Performs linear prediction by autocorrelation on 50ms windows;
5. Root solves the linear prediction polynomial to obtain spectral peaks;

6. Finds the best assignment of peaks to formants over each voiced region of the signal using a dynamic programming algorithm.

Example output is shown in Figure D.3.2.

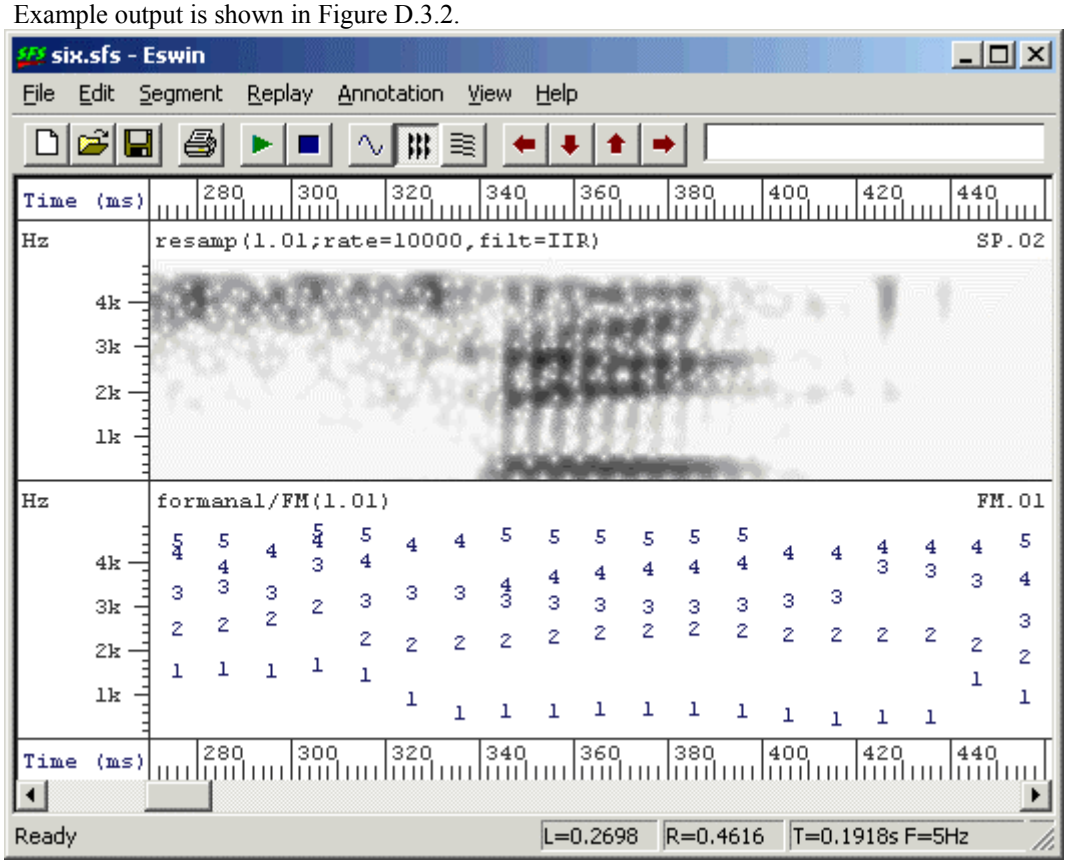


Figure D.3.2 - Example of formant estimation output

Pitch-synchronous analysis

Fixed-frame analysis works well in many situations, and you should certainly try the formanal program on your recordings before attempting anything more sophisticated.

However, with its large windows and DP tracking, formanal will tend to produce rather smooth formant contours which may not reflect accurately the moment-to-moment changes of the vocal tract. A potentially more exact means of obtaining formant frequency measurements is to analyse the data *pitch-synchronously*. Pitch synchronous formant analysis divides the signal up into windows according to a set of pitch epoch markers, such that each analysis window is simply one pitch period long. The result is a set of formant estimates that are output at a rate of one frame per pitch period rather than one frame per 10ms. There are other technical reasons why we expect individual pitch periods as being a better basis for formant estimation.

To perform pitch-synchronous formant analysis, we can use the SFS fmanal program. This program is less sophisticated than formanal and we have to do some careful preparation of the signal before running it. In particular we need to downsample the signal to about 10,000 samples/sec and we need to generate a set of pitch epoch markers. We will discuss these in turn:

Downsampling

If the signal is sampled at a rate higher than about 12,000 samples/sec, it is suggested that you first downsample the signal to about 10,000 samples/sec. To do this, select the signal in SFSWin and choose Tools|Speech|Process|Resample, see Figure D.3.3. Put in a sampling rate of 10,000 samples/sec (or if the original signal is at 22050 or 44100 samples/sec, put in 11025 samples/sec).

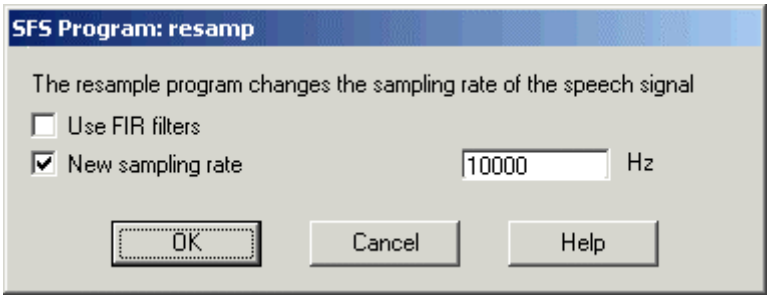


Figure D.3.3 - SFSWin resampling dialog

Pitch epoch marking if Laryngograph signal available

The most reliable way to obtain pitch epoch markers is to make a Laryngograph recording at the same time as the speech signal is recorded. The Laryngograph (www.laryngograph.com) is a specialist piece of equipment that uses two neck electrodes to monitor vocal fold contact area. The resulting stereo signal can be recorded directly into SFSWin or imported from a file using Item|Import|Lx.

From the Laryngograph signal, a set of pitch epoch markers (Tx) can be found from Tools|Lx|Pitch period estimation. Laryngograph recordings are not available for the sample of speech described in Appendix A because we did not want to run the remote risk that attaching the electrodes affected the speech samples obtained. Future research on stammering may use laryngograph signals and allow a convenient way of performing pitch synchronous analysis.

Pitch epoch marking without a Laryngograph signal

A less reliable means to get pitch epoch markers is to analyse the speech signal for periodicity. This can work well for clean, non-reverberant audio signals. To do this, select the speech signal in SFSWin and choose Tools|Speech|Analysis|Fundamental frequency|Pitch epoch location and track (see Figure D.3.4).

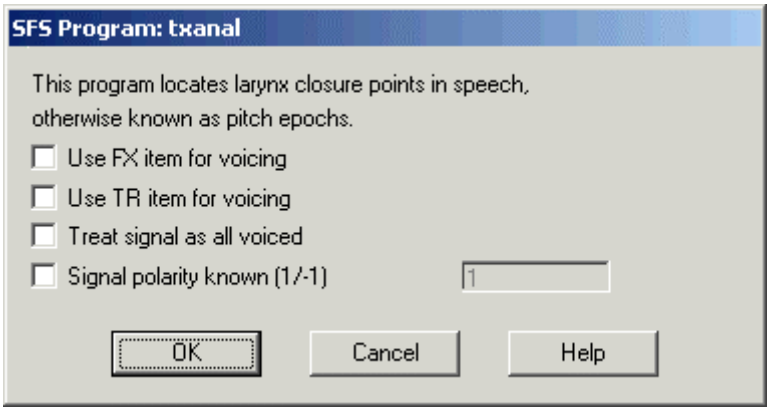


Figure D.3.4 - SFSWin pitch epoch track dialog

The result of the preparation should be two items in the SFS file: a downsampled speech signal and a set of pitch epoch markers (Tx). To perform pitch-synchronous formant analysis, select these two items and choose Tools|Speech|Analysis|Formants estimation. Then select the option "Use Tx for pitch synchronous at offset", leaving the offset value as the default of 0. See Figure D.3.5.

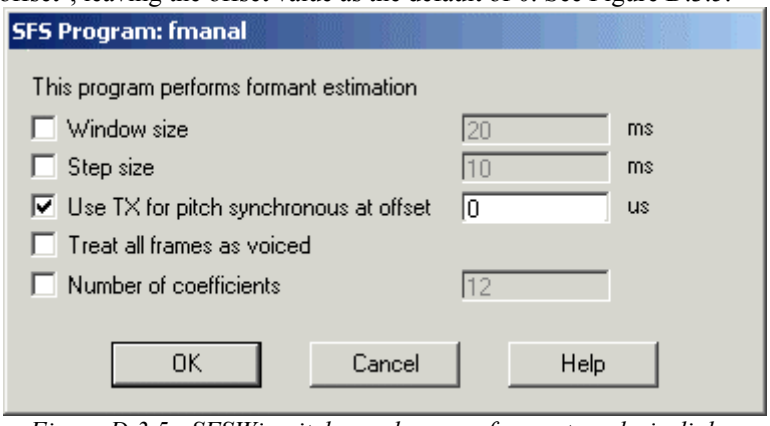


Figure D.3.5 - SFSWin pitch-synchronous formant analysis dialog

An example of pitch synchronous formant analysis is shown in Figure D.3.6. You can see that the formant frames now occur once per pitch period rather than once per 10ms.

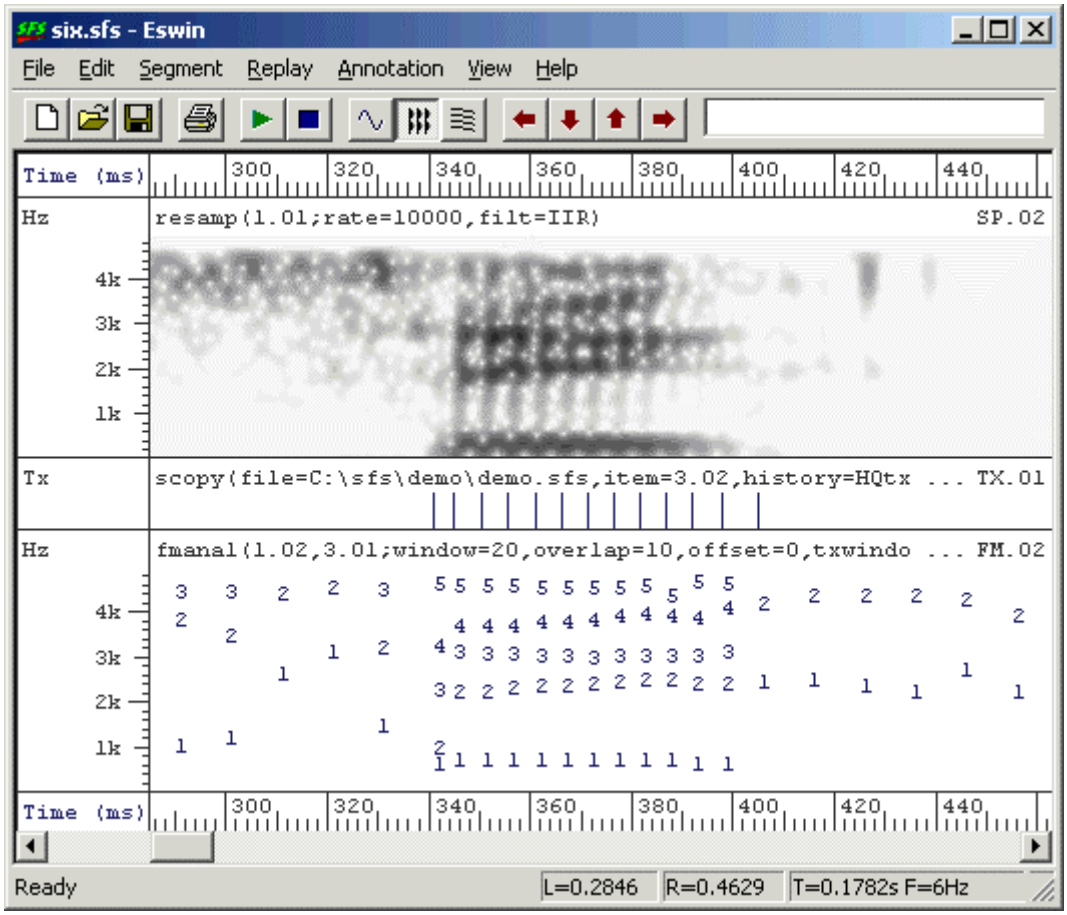


Figure D.3.6 - Example of pitch-synchronous formant estimation output

4. Finding average formant frequencies

We are now in a position to find average formant frequencies in our data. There are three types of average we could consider: (i) the average within a vowel segment, (ii) the average over all segments of a given type for one speaker, or (iii) the average over all vowel segments of a given type spoken by multiple speakers. We will look at these in turn.

Average within a segment

We have annotated our speech signal with labels identifying the vocalic regions where we would like to make a single formant frequency measurement. However, within that region the formant analysis program may have delivered a number of frames of formant estimates. Also the region may cover the whole vocalic segment, while we are interested in a single value which "characterises" the vowel segment. Thus we need to decide how to calculate a characteristic value over what part of the annotated region. In the process we need to take into account the typical contextual changes that occur to vowel formant frequencies in syllables and the typical errors made by formant frequency estimation techniques.

Method 1. Mean over whole segment

We'll start with the most obvious: taking a mean over the whole segment. To demonstrate this, we'll write a script to extract the mean F1, F2 and F3 of each annotated region and save these to a text file in "comma-separated value" (CSV) format. The script is as follows:

```
/* fmsummary.sml - summarise formant measurements from labels */
```

```
/* get mean formant value for a segment */
function var measure_mean(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var    sum;    /* sum of values */
    var    cnt;    /* # values */

    /* calculate mean over whole segment */
    sum=0;
    cnt=0;
    t=next(FM,stime);
    while (t < etime) {
        sum = sum + fm(pno,t);
        cnt = cnt + 1;
        t = next(FM,t);
    }

    return(sum/cnt);
}

/* for each file to be processed */
main {
    var    num;    /* # annotated regions */
    var    i;
    var stime,etime;
    var    vf1,vf2,vf3;

    num = numberof(".");
    /* for each annotation */
    for (i=1;i<=num;i=i+1) if (compare(matchn(".",i),"/")!=0) {
        stime = timen(".",i);
        etime = stime + lengthn(".",i);
        vf1 = measure_mean(stime,etime,5);
        vf2 = measure_mean(stime,etime,8);
        vf3 = measure_mean(stime,etime,11);
        /* output in CSV format */
        print "\"",$filename,"\",\"",matchn(".",i),"\","
        print vf1,",",vf2,",",vf3,"\n"
    }
}
```

This script calls a function measure_mean() for each annotated region for each formant parameter. The script assumes that there is already a FORMANT item in the file, which can be either fixed-frame or pitch synchronous. To run this script, select the annotation item and the formant item to be processed and choose menu option Tools|Run SML script. Enter the file containing the script above and the name of a text file to receive the output, see Figure D.4.1.

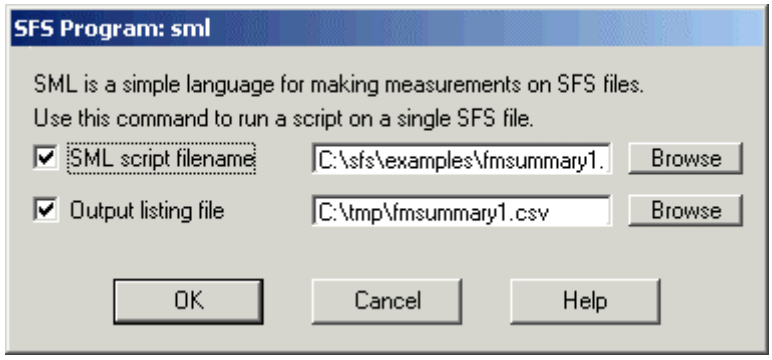


Figure D.4.1 - SFSWin run SMl script dialog

The result of running this script looks like this:

```
"C:\data\ABI\short\brm_f_01_01.sfs","sil", 1120.2540, 2913.2937,
4104.5054
"C:\data\ABI\short\brm_f_01_01.sfs","k", 1146.8917, 2651.2722,
3779.6567
"C:\data\ABI\short\brm_f_01_01.sfs","ae",  789.7859, 1805.1616,
2465.6825
"C:\data\ABI\short\brm_f_01_01.sfs","ng",  662.5625, 1246.0687,
2189.2784
"C:\data\ABI\short\brm_f_01_01.sfs","g",  395.1142,  918.3024,
2061.2703
"C:\data\ABI\short\brm_f_01_01.sfs","ax",  375.5445, 1579.8998,
2311.1436
"C:\data\ABI\short\brm_f_01_01.sfs","r",  438.3379, 1350.3599,
2410.2007
"C:\data\ABI\short\brm_f_01_01.sfs","uw",  462.2808, 1760.6830,
2486.4146
...
```

This CSV format is convenient to use because many spreadsheet and statistics packages can read data in this format. Figure D.4.2 shows this data loaded into Excel.

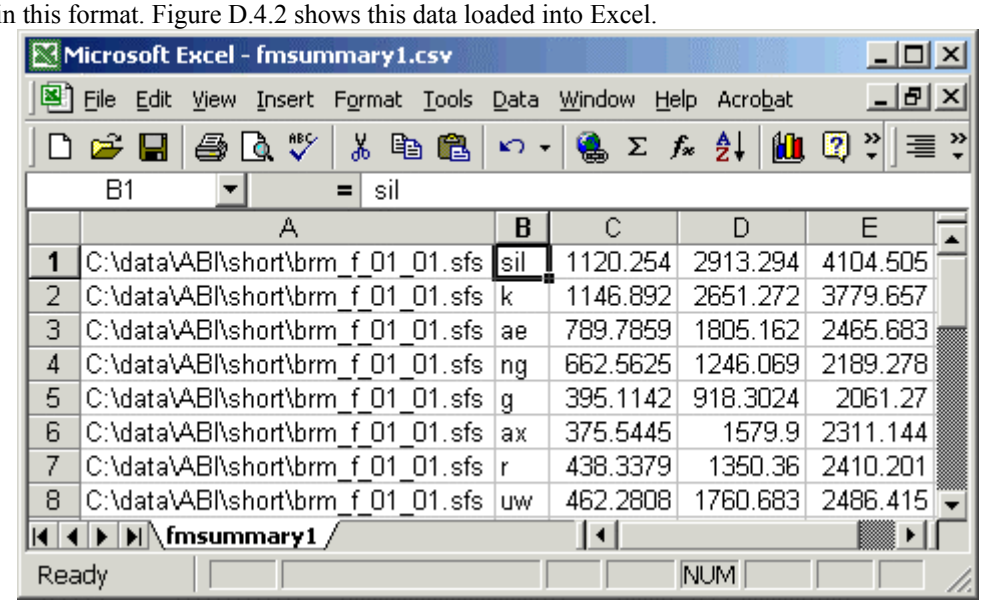


Figure D.4.2 - Formant data loaded into Excel

Method 2. Mean over middle third of segment

Since we expect formant values at the edges of the segment to be less characteristic of the segment than values towards the middle, a refinement of method 1 would be to restrict the analysis to the central third of the segment in time. Here is a replacement measurement function for the script above:

```
/* get mean formant value for a segment */
function var measure_mean_third(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var    sum;    /* sum of values */
    var    cnt;    /* # values */
    var    len;

    /* adjust times to central third */
    len = etime - stime;
    stime = stime + len/3;
    etime = etime - len/3;

    /* calculate mean */
    sum=0;
    cnt=0;
    t=next(FM,stime);
    while (t < etime) {
        sum = sum + fm(pno,t);
        cnt = cnt + 1;
        t = next(FM,t);
    }

    return(sum/cnt);
}
```

Method 3. Median over whole segment

The disadvantage of the mean is that we know that formant tracking errors can occasionally produce wildly inaccurate frequency values. For example, a common tracking error is to relabel F2 as F1, and F3 as F2, and so on. It would seem to be a good idea to remove from the calculation any outlier values. One easy way to do this is to calculate the median over the segment rather than the mean. Here is the adjustment to our script:

```
/* calculate a median */
function var median(table,len)
{
    var table[];    /* array of values */
    var len;        /* # values */
    var    i,j,tmp;

    /* sort table */
    for (i=2;i<=len;i=i+1) {
        j = i;
        tmp = table[j];
        while (table[j-1] > tmp) {
            table[j] = table[j-1];
            j = j - 1;
            if (j==1) break;
        }
        table[j] = tmp;
    }

    /* return middle value */
    if ((len%2)==1) {
        return(table[1+len/2])
    }
    else {
        return((table[len/2]+table[1+len/2])/2);
```

```
    }
}

/* get median formant value for a segment */
function var measure_median(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */

    /* calculate median */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(median(af,nf));
}
```

Method 4. Trimmed mean over whole segment

The disadvantage of the median is that it only picks one value as representative of the formant contour for the segment. One way to get a smoothed estimate but disregard outliers is to use the "trimmed mean" - that is the mean of the values at the middle of the distribution. In the following variation we calculate a trimmed mean of the central 60% (disregarding the lowest 20% and the highest 20%) - but adjust as you see fit.

```
/* calculate trimmed mean */
function var trimmean(table,len)
{
    var table[];    /* array of values */
    var len;        /* # values */
    var    I,j,tmp;
    var    lo,hi;

    /* sort table */
    for (i=2;i<=len;i=i+1) {
        j = i;
        tmp = table[j];
        while (table[j-1] > tmp) {
            table[j] = table[j-1];
            j = j - 1;
            if (j==1) break;
        }
        table[j] = tmp;
    }

    /* find mean over middle portion */
    lo = trunc(0.5 + 1 + len/5);    /* lose bottom 20% */
    hi = trunc(0.5 + len - len/5);    /* lose top 20% */
    j=0;
    tmp=0;
    for (i=lo;i<=hi;i=i+1) {
        tmp = tmp + table[i];
        j = j + 1;
```

```
    }
    return(tmp/j);
}

/* get median formant value for a segment */
function var measure_trimmed_mean(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */

    /* calculate trimmed mean */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(trimmean(af,nf));
}
```

Method 5. Find straight line of best fit

Since we don't expect the formant frequency to be constant over the segment, another approach is to fit a line to the formant values and choose the value of that line at the centre point of the segment as representative of the segment as a whole. To fit a line, we perform a least-squares procedure as follows:

```
/* calculate last-squares fit and return value at time */
function var lsqfit(at,af,nf,t)
{
    var    at[];    /* array of times */
    var    af[];    /* array of frequencies */
    var    nf;        /* # values */
    var    t;        /* output time */
    var    I
    stat x,y,xy
    var    a,b;    /* coefficients */

    /* collect parameters */
    for (i=1;i<=nf;i=i+1) {
        x += at[i];
        y += af[i]
        xy += at[i]*af[i]
    }

    /* find coefficients of line */
    b = (nf*xy.sum-x.sum*y.sum)/(nf*x.sumsq-x.sum*x.sum);
    a = (y.sum - b*x.sum)/nf;

    return(a + b*t);
}

/* get mid point of formant trajectory for a segment */
function var measure_linear(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
```

```
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var at[1:1000];    /* array of time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */

    /* calculate trajectory */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        at[nf] = t;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(lsqfit(at,af,nf,(stime+etime)/2));
}
```

Method 6. Find quadratic of best fit

Finally, we refine the last approach by fitting a quadratic rather than a straight line to the formant values. This accommodates the fact that formant trajectories are often curved through a segment. A possible disadvantage is that we may become over sensitive to outliers. Here are the modifications needed:

```
/* calculate last-squares fit of quadratic and return value at time */
function var quadfit(at,af,nf,t)
{
    var    at[];    /* array of times */
    var    af[];    /* array of frequencies */
    var    nf;        /* # values */
    var    t;        /* output time */
    var    i;
    var    a,b,c;    /* coefficients */
    var    mat1[1:4];    /* normal matrix row 1 */
    var    mat2[1:4];    /* normal matrix row 2 */
    var    mat3[1:4];    /* normal matrix row 3 */

    /* collect parameters */
    for (i=1;i<=nf;i=i+1) {
        mat1[1] = mat1[1] + 1;
        mat1[2] = mat1[2] + at[i];
        mat1[3] = mat1[3] + at[i] * at[i];
        mat1[4] = mat1[4] + af[i];
        mat2[1] = mat2[1] + at[i];
        mat2[2] = mat2[2] + at[i] * at[i];
        mat2[3] = mat2[3] + at[i] * at[i] * at[i];
        mat2[4] = mat2[4] + at[i] * af[i];
        mat3[1] = mat3[1] + at[i] * at[i];
        mat3[2] = mat3[2] + at[i] * at[i] * at[i];
        mat3[3] = mat3[3] + at[i] * at[i] * at[i] * at[i];
        mat3[4] = mat3[4] + at[i] * at[i] * af[i];
    }

    /* reduce lines 2 and 3, column 1 */
    for (i=4;i>=1;i=i-1) {
        mat2[i] = mat2[i] - mat1[i]*mat2[1]/mat1[1];
        mat3[i] = mat3[i] - mat1[i]*mat3[1]/mat1[1];
    }

    /* reduce line 3 column 2 */
    for (i=4;i>=2;i=i-1) {
```

```
        mat3[i] = mat3[i] - mat2[i]*mat3[2]/mat2[2];
    }

    /* calculate c */
    c = mat3[4]/mat3[3];

    /* back substitute to get b */
    b = (mat2[4] - mat2[3]*c)/mat2[2];

    /* back substitute to get a */
    a = (mat1[4] - mat1[3]*c - mat1[2]*b)/mat1[1];

    return(a + b*t + c*t*t);
}

/* get mid point of formant trajectory for a segment */
function var measure_quadratic(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var at[1:1000];    /* array of time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */

    /* calculate trajectory */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        at[nf] = t;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(quadfit(at,af,nf,(stime+etime)/2));
}
```

In the next section we will apply these approaches to the study of the distribution of formant values for a particular segment type for a single speaker, and investigate which gives us the least variable results.

Average within a speaker

So far we have shown a number of ways in which to extract a characteristic formant frequency value from each annotated region of the signal. In this section we will look at the distribution of those values for a number of instances of a single type of annotated region for a single speaker. This will not only demonstrate how to collect data across a number of files, but it will also allow us to make a simple empirical study of the performance of the six different methods. Attention aimed at obtaining the most accurate method may reap particular benefits in disordered speech where the formant values estimated by different methods may be more variable than with fluent speakers.

The script below calls the measure_mean() function on all instances of a given labeled region found in the input files. It then collects the values into a histogram and plots the histogram and a modelled normal distribution for each formant. It also reports the mean and standard deviation of the estimated characteristic formant frequencies for the segment.

```
/* fmplot1.sml -- plot distribution of formant frequency averages */
/* - uses mean over whole annotated region */
```

```
stat    f1;             /* f1 distribution */
stat    f2;             /* f2 distribution */
stat    f3;             /* f3 distribution */
var        hf1[0:100];     /* f1 histogram (50Hz bins) */
var        hf2[0:100];     /* f2 histogram (50Hz bins) */
var        hf3[0:100];     /* f3 histogram (50Hz bins) */

string    label;    /* annotation label to measure */
file    gop;    /* graphics output */

/* get mean formant value for a segment */
function var measure_mean(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var    sum;    /* sum of values */
    var    cnt;    /* # values */

    /* calculate mean over whole segment */
    sum=0;
    cnt=0;
    t=next(FM,stime);
    while (t < etime) {
        sum = sum + fm(pno,t);
        cnt = cnt + 1;
        t = next(FM,t);
    }

    if (cnt > 0) return(sum/cnt) else return(ERROR);
}

/* normal distribution */
function var normal(st,x)
stat st;
{
    var x;
    x = x - st.mean;
    return(exp(-0.5*x*x/st.variance)/sqrt(2*3.14159*st.variance));
}

/* plot histogram overlaid with normal distribution */
function var plotdist(st,hs)
stat st;
var hs[];
{
    var    i;
    var    xdata[1:2];
    var ydata[0:4000];

    /* set up x-axes */
    xdata[1]=0;
    xdata[2]=4000;
    plotxdata(xdata,1)

    /* plot histogram */
    plotparam("type=hist");
    for (i=0;i<=80;i=i+1) ydata[i] = hs[i]/st.count;
    plot(gop,1,ydata,81);
```

```
    /* plot normal distribution */
    plotparam("type=line");
    for (i=0;i<4000;i=i+1) ydata[i]=50*normal(st,i);
    plot(gop,1,ydata,4000);
}

/* record details of a single segment */
function var recordsegment(stime,etime,pno,st,hs)
stat st;
var hs[];
{
    var stime,etime,pno;
    var f

    f = measure_mean(stime,etime,pno);
    if (f) {
        st += f;
        hs[trunc(f/50)] = hs[trunc(f/50)] + 1;
    }
}

/* initialise */
init {
    string ans

    /* get label */
    print#stderr "Enter label to find : ";
    input label;

    /* where to send graphs */
    print#stderr "Send graph to file 'dig.gif' ? (Y/N) "
    input ans
    if (index("^[yY]",ans)) {
        openout(gop,"|dig -g -s 500x375 -o dig.gif");
    } else openout(gop,"|dig");
}

/* for each file to be processed */
main {
    var        i,num,stime,etime

    num=numberof(label);
    print#stderr "File ",$filename," has ",num," matching
annotations\n";

    for (i=1;i<=num;i=i+1) {
        stime = next(FM,timen(label,i));
        etime = timen(label,i) + lengthn(label,i);
        recordsegment(stime,etime,5,f1,hf1);    /* 5 = F1 */
        recordsegment(stime,etime,8,f2,hf2);    /* 8 = F2 */
        recordsegment(stime,etime,11,f3,hf3);    /* 11 = F3 */
    }
}

/* display summary statistics and graphs */
summary {
    print#stderr "F1 = ",f1.mean," +/-",f1.stddev,"Hz
(",f1.count:1,")\n";
    print#stderr "F2 = ",f2.mean," +/-",f2.stddev,"Hz
(",f2.count:1,")\n";
    print#stderr "F3 = ",f3.mean," +/-",f3.stddev,"Hz
```

```
(",f3.count:1,")\n";

    plottitle(gop,"/"++label++"/ formant distributions 1");
    plotparam("title=(mean over whole segment)");
    plotparam("xtitle=Frequency (Hz)");
    plotparam("ytitle=Probability");

    if (f1.variance > 0) plotdist(f1,hf1);
    if (f2.variance > 0) plotdist(f2,hf2);
    if (f3.variance > 0) plotdist(f3,hf3);
}
```

We will run this script over 200 phonetically annotated sentences that form part of the SCRIBE corpus. We will just look at the distribution of the formant frequencies among 65 instances of /A:/ in those sentences. The graphical output of the script is shown in Figure D.4.3.

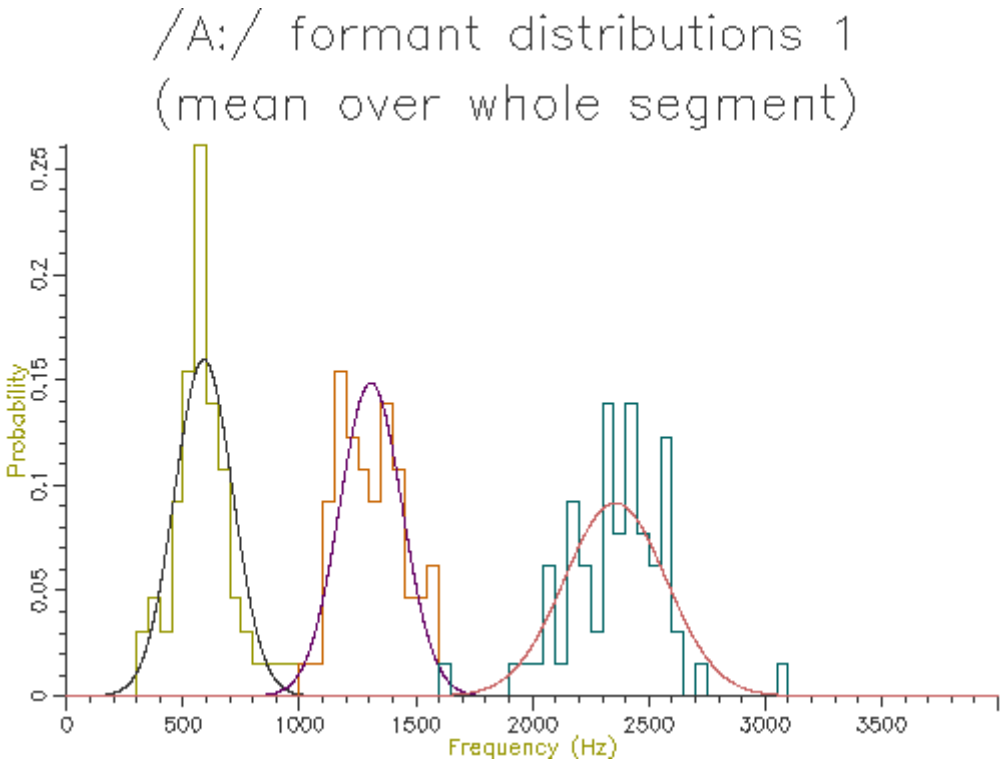


Figure D.4.3 - Analysis of 65 /A:/ vowels, method 1

This figure shows quite clearly the breadth of the formant frequency distributions even when all the vowels are from the same speaker. The estimated characteristic formant frequencies for /A:/ for this speaker are also output by the script:

F1 = 579.2066 +/- 112.6763Hz (65)
F2 = 1278.4254 +/- 127.9527Hz (65)
F3 = 2332.7904 +/- 210.0033Hz (65)

Figures D.4.4 to D.4.8 show the output of similar scripts set up to use each of the other methods described in the last section

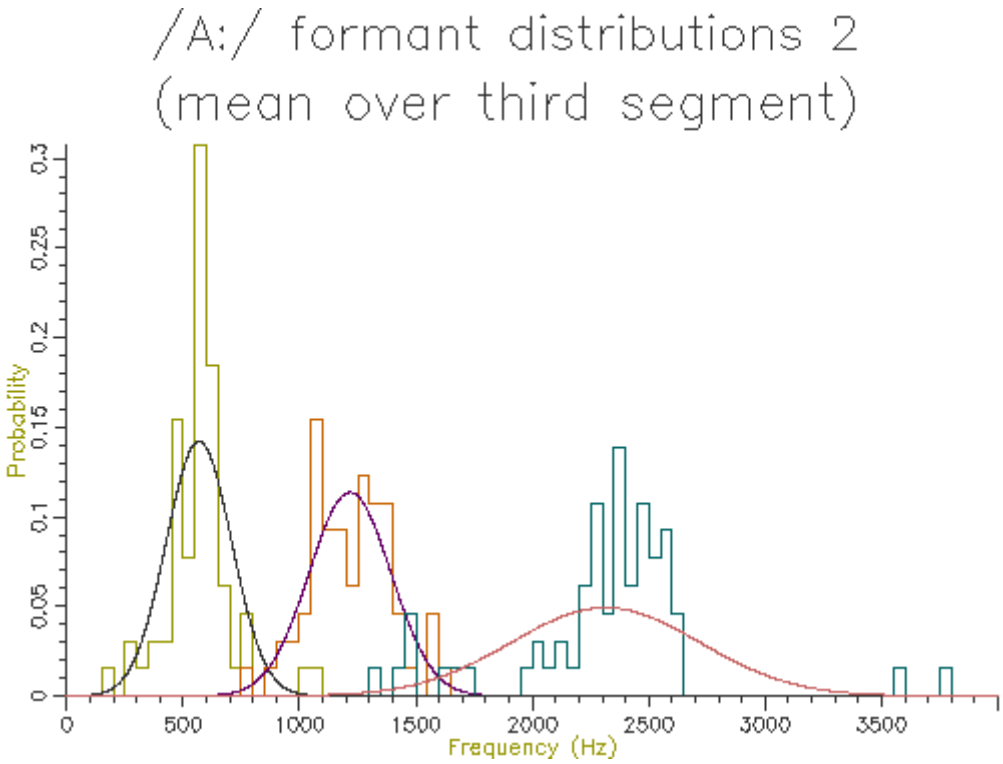


Figure D.4.4 - Analysis of 65 /A:/ vowels, method 2

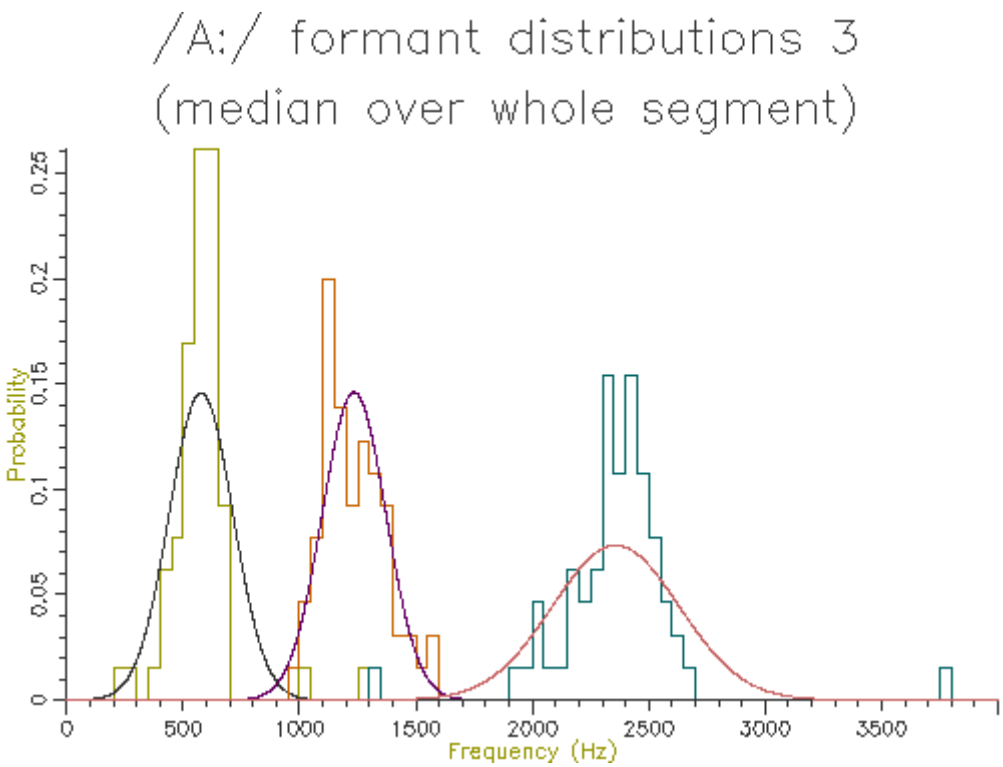


Figure D.4.5 - Analysis of 65 /A:/ vowels, method 3

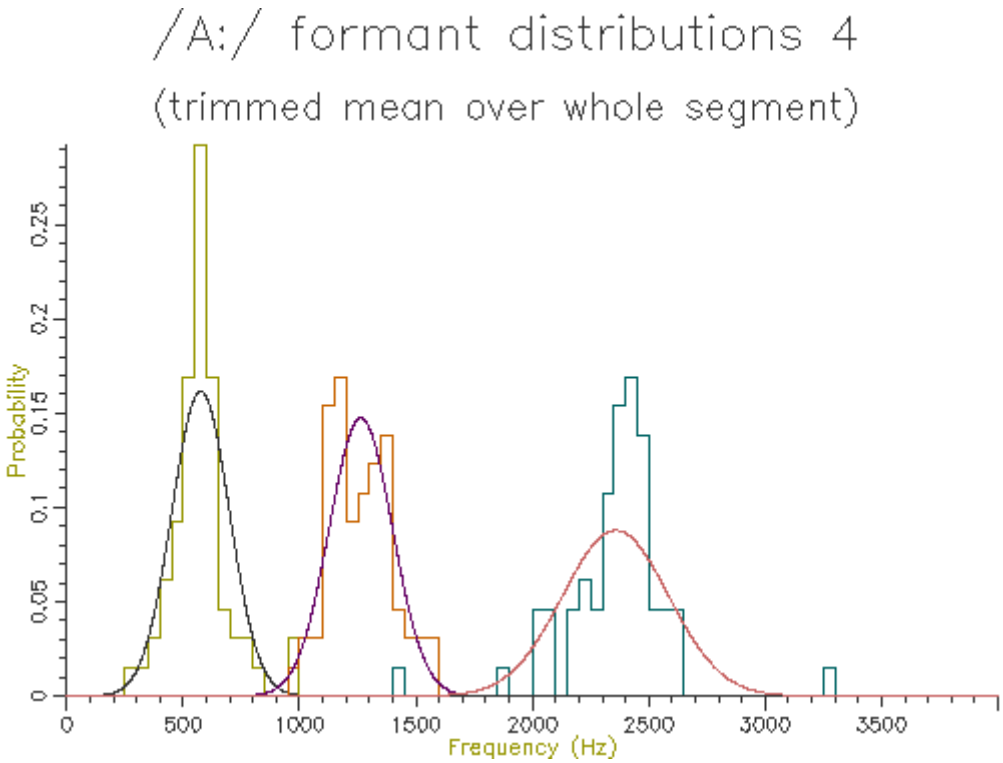


Figure D.4.6 - Analysis of 65 /A:/ vowels, method 4

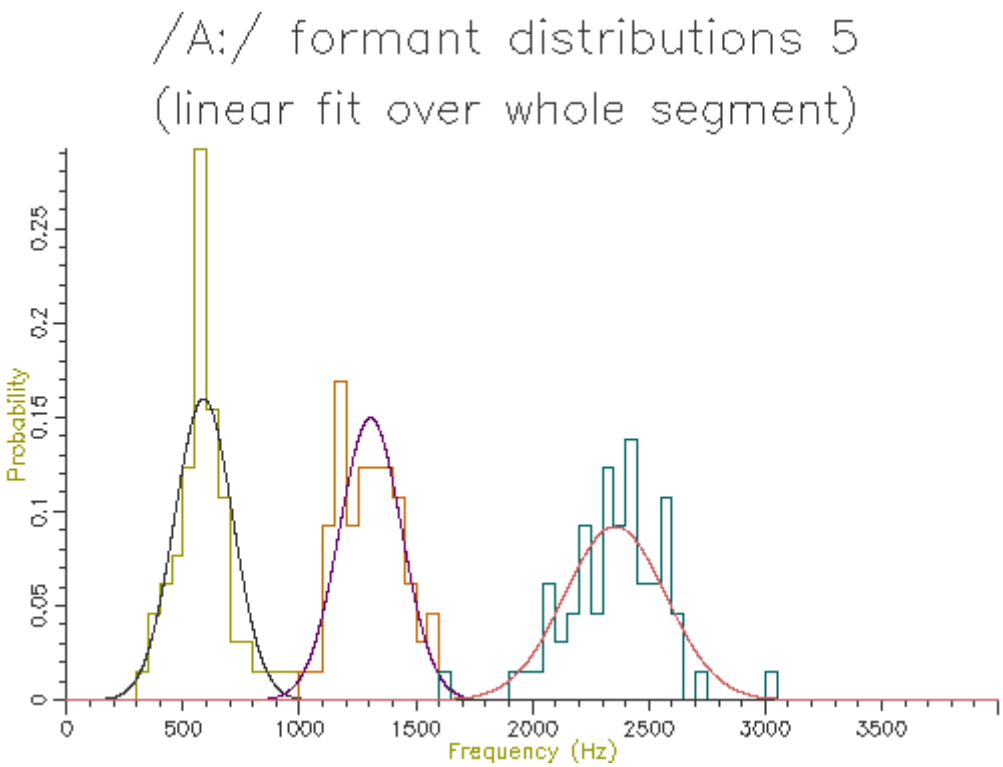


Figure D.4.7 - Analysis of 65 /A:/ vowels, method 5

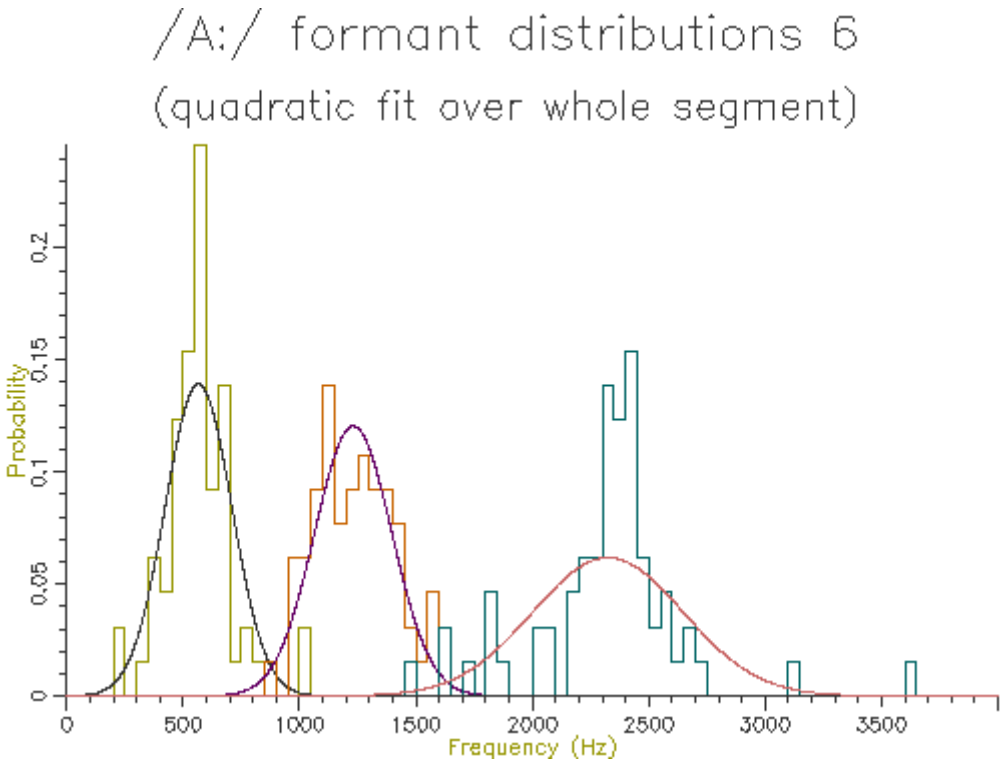


Figure D.4.8 - Analysis of 65 /A:/ vowels, method 6

The table below shows the mean and standard deviation of the characteristic formant frequency for each formant for each of the analysis methods:

Method	F1		F2		F3	
	Mean (Hz)	Dev (Hz)	Mean (Hz)	Dev (Hz)	Mean (Hz)	Dev (Hz)
1. Mean whole	588.6	125.1	1306.2	134.3	2356.0	217.7
2. Mean third	567.8	140.4	1216.6	175.6	2314.1	404.6
3. Median	576.9	136.9	1234.2	136.7	2357.8	272.2
4. Trimmed mean	575.3	123.4	1263.0	135.3	2357.2	227.2
5. Fit line	588.2	125.0	1305.0	133.6	2355.1	217.0
6. Fit quadratic	565.9	143.3	1230.5	165.6	2324.0	323.1

Looking at the table, the lowest variance for F1 comes through using the trimmed mean, while the lowest variance for F2 and F3 comes from the straight line fit. Which is the best method? The problem is choosing a method that is robust to the typical errors in formant estimation. The trimmed mean seems a simple and robust measure (at least for /A:/).

When studying stammered speech, the differences may be subtle, so care should be taken to use the most accurate procedure. The next section considers requirements for representing audio data across groups of speakers.

Speaker Normalisation

So far we have looked at collecting formant measurements within a segment and across copies of a segment within one speaker. The data in Appendix A come from speakers who are heterogeneous in gender, age and accent. In this section we look at the problems of collecting formant measurements across speakers. The biggest challenge we face is the standardisation of the range of formant values for each speaker prior to averaging across speakers. Because speakers are of physically different sizes, the

absolute value of their formant frequencies will vary because of their size as well as because of any change in accent or style.

We will only look at a simple means for standardising or normalising formant frequencies. As well as collecting formant measurements from a collection of recordings of a speaker for a single segment type, we will also collect measurements for all related types for the speaker. We can then represent the characteristic formant frequencies for a segment for a speaker in terms of their relationship to the overall distribution of frequencies for the speaker.

To demonstrate the idea we will first show how to collect segment specific and general measurements for vowels from a number of annotated recordings of a single speaker, delivering a normalised formant estimate. We will use the trimmed mean to get a characteristic value for each segment.

```
/* fmnorm.sml - calculate normalised formant measurements for segment
*/

/* global distribution */
stat    gf1,gf2,gf3;

/* segment specific distribution */
stat    sf1,sf2,sf3;

/* label to find */
string    label;

/* calculate trimmed mean */
function var trimmean(table,len)
{
    var table[];    /* array of values */
    var len;        /* # values */
    var    i,j,tmp;
    var    lo,hi;

    /* sort table */
    for (i=2;i<=len;i=i+1) {
        j = i;
        tmp = table[j];
        while (table[j-1] > tmp) {
            table[j] = table[j-1];
            j = j - 1;
            if (j==1) break;
        }
        table[j] = tmp;
    }

    /* find mean over middle portion */
    lo = trunc(0.5 + 1 + len/5);    /* lose bottom 20% */
    hi = trunc(0.5 + len - len/5);    /* lose top 20% */
    j=0;
    tmp=0;
    for (i=lo;i<=hi;i=i+1) {
        tmp = tmp + table[i];
        j = j + 1;
    }
    if (j > 0) return(tmp/j) else return(ERROR);
}

/* get trimmed mean formant value for a segment */
function var measure_trimmed_mean(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
```

```
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */

    /* calculate trimmed mean */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(trimmean(af,nf));
}

/* record details of a single segment */
function var recordsegment(stime,etime,pno,st)
stat st;
{
    var stime,etime,pno;
    var f

    f = measure_trimmed_mean(stime,etime,pno);
    if (f) st += f;
}

/* initialise */
init {
    /* get label */
    print#stderr "Enter label to find : ";
    input label;
}

/* for each file to be processed */
main {
    var    num;    /* # annotated regions */
    var    i;
    var stime,etime;

    num = numberof(".");
    print#stderr "File ",$filename," has ",num," annotations\n";

    /* for each annotation */
    for (i=1;i<=num;i=i+1) {
        stime = timen(".",i);
        etime = stime + lengthn(".",i);
        if (index(label,matchn(".",i))) {
            /* matching label */
            recordsegment(stime,etime,5,sf1);
            recordsegment(stime,etime,8,sf2);
            recordsegment(stime,etime,11,sf3);
        }
        if (index("^[aeiouAEIOU3@&]",matchn(".",i))) {
            /* some kind of vowel */
            recordsegment(stime,etime,5,gf1);
            recordsegment(stime,etime,8,gf2);
            recordsegment(stime,etime,11,gf3);
        }
    }
```

```
}

/* summarise */
summary {
    /* report speaker means */
    print "Speaker means: (",gf1.count:1," segments)\n";
    print "F1 = ",gf1.mean," +/-",gf1.stddev,"Hz\n";
    print "F2 = ",gf2.mean," +/-",gf2.stddev,"Hz\n";
    print "F3 = ",gf3.mean," +/-",gf3.stddev,"Hz\n";

    /* report segment means */
    print "Segment means: (",sf1.count:1," segments)\n";
    print "F1 = ",sf1.mean," +/-",sf1.stddev,"Hz\n";
    print "F2 = ",sf2.mean," +/-",sf2.stddev,"Hz\n";
    print "F3 = ",sf3.mean," +/-",sf3.stddev,"Hz\n";

    /* report normalised means */
    print "Normalised segment means: (",sf1.count:1," segments)\n";
    print "F1 = ",(sf1.mean-gf1.mean)/gf1.stddev," z-score\n";
    print "F2 = ",(sf2.mean-gf2.mean)/gf2.stddev," z-score\n";
    print "F3 = ",(sf3.mean-gf3.mean)/gf3.stddev," z-score\n";
}
```

In this script we collect the mean value of F1, F2 and F3 for a single segment type, and also collect the mean and variance of F1, F2 and F3 over all vowel segments. We then express F1, F2 and F3 for the given segment as *z-score* positions of the segment mean with respect to the mean and variance of all vowels. The table below shows the output of the script over 200 sentences spoken by one person for two different vowels:

/A:/ vowel

```
Speaker means: (1847 segments)
F1 =   426.0644 +/-  151.5595Hz
F2 =  1589.4008 +/-  314.8747Hz
F3 =  2496.9879 +/-  239.9074Hz
Segment means: (65 segments)
F1 =   575.2579 +/-  123.4225Hz
F2 =  1263.0107 +/-  135.3098Hz
F3 =  2357.1892 +/-  227.2396Hz
Normalised segment means: (65 segments)
F1 =     0.9844 z-score
F2 =    -1.0366 z-score
F3 =    -0.5827 z-score
```

/i:/ vowel

```
Speaker means: (1847 segments)
F1 =   426.0644 +/-  151.5595Hz
F2 =  1589.4008 +/-  314.8747Hz
F3 =  2496.9879 +/-  239.9074Hz
Segment means: (192 segments)
F1 =   336.3615 +/-   83.0778Hz
F2 =  1936.3488 +/-  211.8316Hz
F3 =  2621.4353 +/-  209.6236Hz
Normalised segment means: (192 segments)
F1 =    -0.5919 z-score
F2 =     1.1019 z-score
F3 =     0.5187 z-score
```

We can now apply this idea to compare formant frequencies across speakers. In this demonstration we will plot the mean F1 and F2 for 5 long monophthongs (/i:/, /u:/, /3:/, /A:/, /O:/) over 10 male and

10 female speakers of a single accent. We will do this first without normalisation, then with normalisation.

Note: in the data used for this demonstration, speakers can be identified from the filename: the first 8 characters of the filename are specific to the speaker. We will use this to collect the F1 and F2 data on a speaker-dependent basis. Also this database is labelled with ARPABET symbols, not JSRU or SAMPA.

```
/* f1f2speaker.sml - F1-F2 diagram for vowels across speakers */

/* list of speakers */
string    stab[1:100];
var    scnt;

/* general vowel distributions */
stat    gf1[1:100];
stat    gf2[1:100];

/* specific vowel distributions */
stat    s1f1[1:100],s1f2[1:100];
stat    s2f1[1:100],s2f2[1:100];
stat    s3f1[1:100],s3f2[1:100];
stat    s4f1[1:100],s4f2[1:100];
stat    s5f1[1:100],s5f2[1:100];

/* output file */
file    gop;

/* function to find speaker code from filename */
function var speakercode(name)
{
    var    code;
    string name;

    name = name:8;    /* first eight characters */
    code = entry(name,stab);
    if (code) return(code);
    scnt=scnt+1;
    stab[scnt]=name;
    return(scnt);
}

/* calculate trimmed mean */
function var trimmean(table,len)
{
    var table[];    /* array of values */
    var len;        /* # values */
    var    i,j,tmp;
    var    lo,hi;

    /* sort table */
    for (i=2;i<=len;i=i+1) {
        j = i;
        tmp = table[j];
        while (table[j-1] > tmp) {
            table[j] = table[j-1];
            j = j - 1;
            if (j==1) break;
        }
        table[j] = tmp;
    }
```

```
    /* find mean over middle portion */
    lo = trunc(0.5 + 1 + len/5);    /* lose bottom 20% */
    hi = trunc(0.5 + len - len/5);    /* lose top 20% */
    j=0;
    tmp=0;
    for (i=lo;i<=hi;i=i+1) {
        tmp = tmp + table[i];
        j = j + 1;
    }
    if (j > 0) return(tmp/j) else return(ERROR);
}

/* get trimmed mean formant value for a segment */
function var measure_trimmed_mean(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */

    /* calculate trimmed mean */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(trimmean(af,nf));
}

/* record details of a single segment */
function var recordsegment(stime,etime,stf1,stf2)
stat stf1;
stat stf2;
{
    var stime,etime;
    var f

    f = measure_trimmed_mean(stime,etime,5);
    if (f) stf1 += f;
    f = measure_trimmed_mean(stime,etime,8);
    if (f) stf2 += f;
}

/* plot F1-F2 for segment */
function var plotf1f2segment(label,astf1,astf2)
{
    string    label;
    stat    astf1[];
    stat    astf2[];
    var    i;
    var    xdata[1:100];
    var    ydata[1:100];

    /* for each speaker */
    for (i=1;i<=scnt;i=i+1) {
        xdata[i] = astf1[i].mean;
```

```
        ydata[i] = astf2[i].mean;
    }
    plotparam("char="++label);

    plotxdata(xdata,0);
    plot(gop,1,ydata,scnt);
}

/* plot F1-F2 graph */
function var plotf1f2()
{
    openout(gop,"|dig");
    plottitle(gop,"Formant variation");
    plotparam("title=no normalisation");
    plotparam("xtitle=F1 Frequency (Hz)");
    plotparam("ytitle=F2 Frequency (Hz)");
    plotparam("type=point");
    plotaxes(gop,1,200,900,1000,2500);

    /* for each segment in turn */
    plotf1f2segment("i",s1f1,s1f2);
    plotf1f2segment("u",s2f1,s2f2);
    plotf1f2segment("3",s3f1,s3f2);
    plotf1f2segment("A",s4f1,s4f2);
    plotf1f2segment("O",s5f1,s5f2);

    close(gop);
}

/* plot F1-F2 for segment */
function var plotf1f2normsegment(label,astf1,astf2)
{
    string    label;
    stat    astf1[];
    stat    astf2[];
    var    i;
    var    xdata[1:100];
    var    ydata[1:100];

    /* for each speaker */
    for (i=1;i<=scnt;i=i+1) {
        xdata[i] = (astf1[i].mean-gf1[i].mean)/gf1[i].stddev;
        ydata[i] = (astf2[i].mean-gf2[i].mean)/gf2[i].stddev;
    }
    plotparam("char="++label);

    plotxdata(xdata,0);
    plot(gop,1,ydata,scnt);
}

/* plot F1-F2 graph */
function var plotf1f2norm()
{
    openout(gop,"|dig");
    plottitle(gop,"Formant variation");
    plotparam("title=with normalisation");
    plotparam("xtitle=F1 Frequency (z-score)");
    plotparam("ytitle=F2 Frequency (z-score)");
    plotparam("type=point");
    plotaxes(gop,1,-1.5,1.5,-2,2);
```

```
    /* for each segment in turn */
    plotf1f2normsegment("i",s1f1,s1f2);
    plotf1f2normsegment("u",s2f1,s2f2);
    plotf1f2normsegment("3",s3f1,s3f2);
    plotf1f2normsegment("A",s4f1,s4f2);
    plotf1f2normsegment("O",s5f1,s5f2);

    close(gop);
}

/* for each file to be processed */
main {
    var    num;    /* # annotated regions */
    var    i;
    var stime,etime;
    var    scode;
    string    label;

    /* get code for speaker */
    scode=speakercode($filename);

    /* report file */
    num = numberof(".");
    print#stderr "File ",$filename," has ",num," annotations\n";

    /* for each annotation */
    for (i=1;i<=num;i=i+1) {
        stime = timen(".",i);
        etime = stime + lengthn(".",i);
        label = matchn(".",i);
        switch (label) {    /* ARPABET labels */
        case "iy": recordsegment(stime,etime,s1f1[scode],s1f2[scode]);
        case "uw": recordsegment(stime,etime,s2f1[scode],s2f2[scode]);
        case "er": recordsegment(stime,etime,s3f1[scode],s3f2[scode]);
        case "aa": recordsegment(stime,etime,s4f1[scode],s4f2[scode]);
        case "ao": recordsegment(stime,etime,s5f1[scode],s5f2[scode]);
        }
        if (index("^[aeiou]",label)) {
            /* some kind of vowel */
            recordsegment(stime,etime,gf1[scode],gf2[scode]);
        }
    }
}

/* summarise collected data */
summary {
    /* plot graph unnormalised */
    plotf1f2();

    /* plot graph normalised */
    plotf1f2norm();
}
```

The outputs of the script on 10 male and 10 female speakers are shown in Figures D.4.9 and D.4.10. The normalised results show considerably less variation across speakers and also less overlap across segment types. In the next section we will look at how we can perform statistical comparisons on these kind of data across speakers and types.

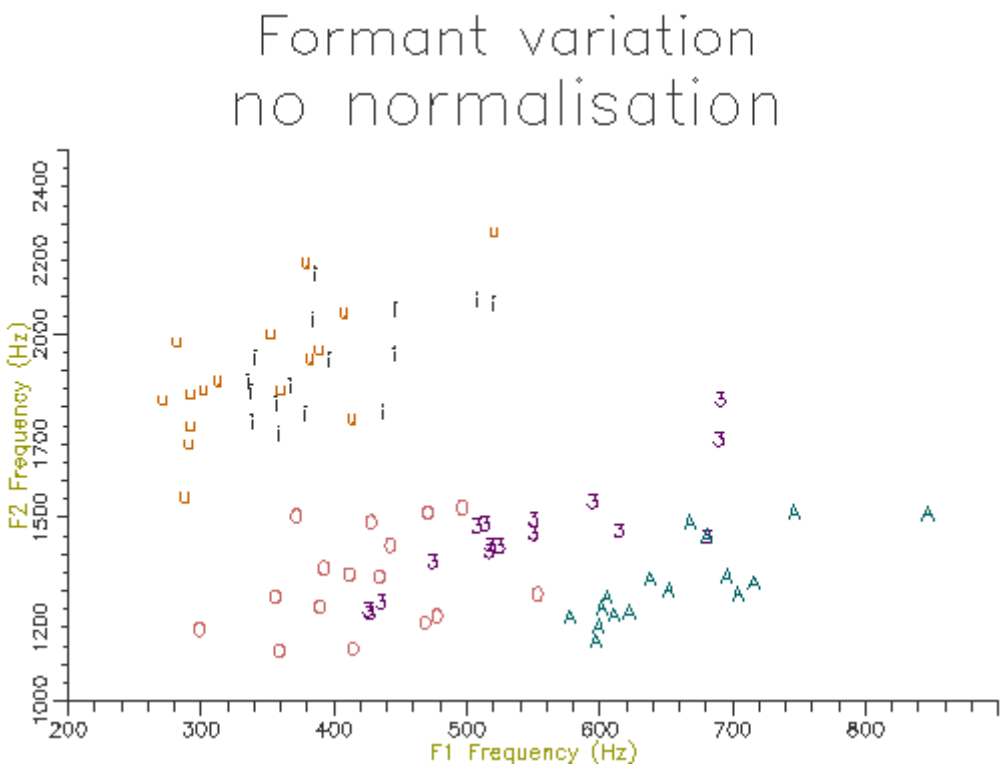


Figure D.4.9 - Vowel analysis without normalization

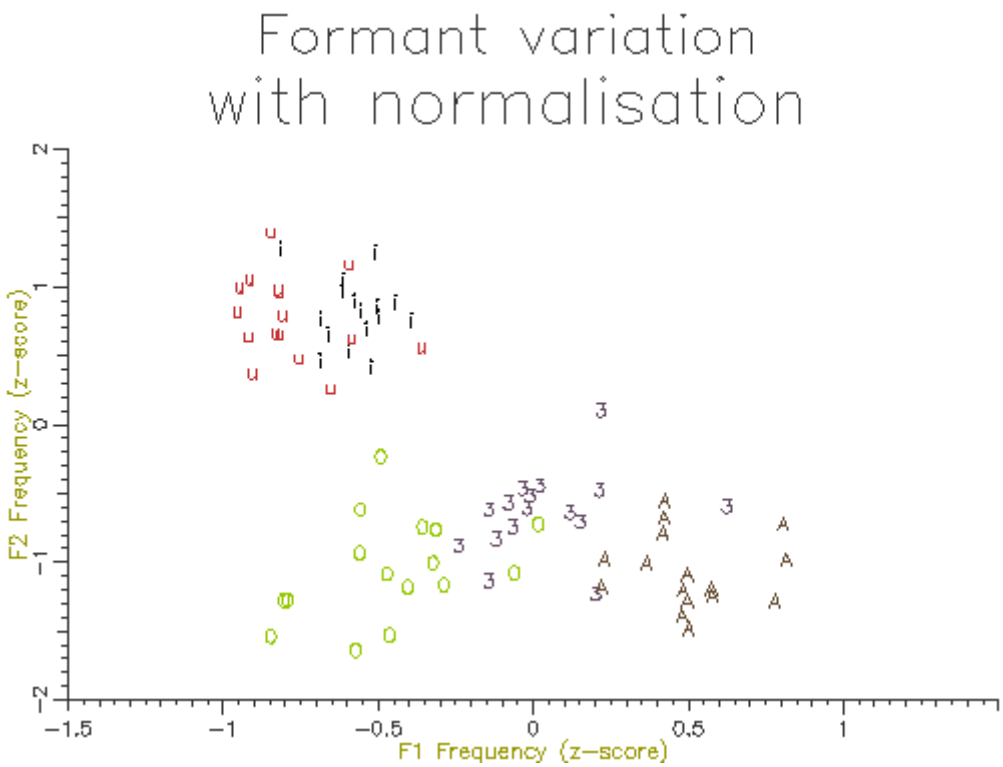


Figure D.4.10 - Vowel analysis with normalization

5. Data Analysis

We have collected together a range of tools for measuring formant frequencies from annotated speech signals. In this section we will look at how we might use those tools to establish the likelihood of various experimental hypotheses using techniques of *inferential statistics*.

+Comparison of means (1D)

As can be seen from the formant frequency distributions shown in Figures D.4.3 to D.4.8, typical formant distributions for a single segment show a fairly normal shape. Thus it is reasonable for us to use a parametric method for comparing the means of two samples. We will show this kind of analysis through a number of worked example cases.

Is a vowel the same in two different contexts?

In this example we look at some instances of /i:/ vowels spoken in word-final and word-medial positions. We can use one of the scripts from section D.4 (e.g. fmsummary4.sml) to extract this data from annotated recordings of a single speaker. Here we have divided it into two sets according to the context in which each vowel occurs:

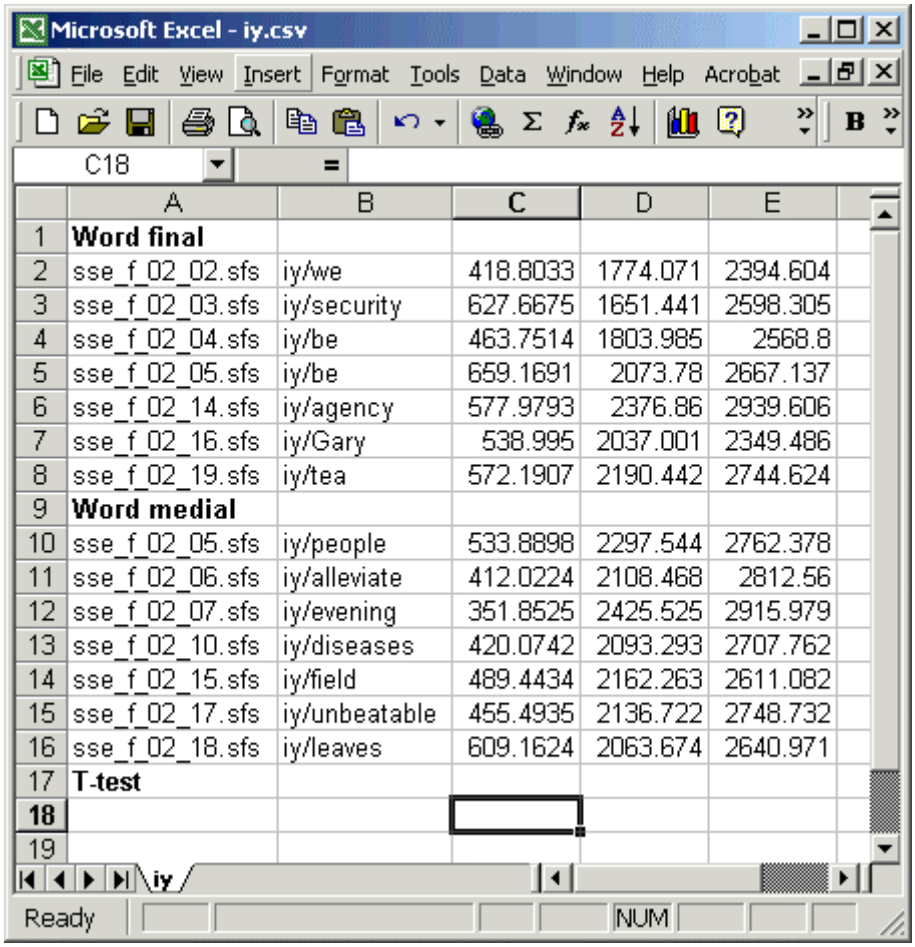
Word final			
"sse_f_02_02.sfs","iy/we",	418.8033,	1774.0710,	2394.6035
"sse_f_02_03.sfs","iy/security",	627.6675,	1651.4409,	2598.3047
"sse_f_02_04.sfs","iy/be",	463.7514,	1803.9850,	2568.7997
"sse_f_02_05.sfs","iy/be",	659.1691,	2073.7795,	2667.1369
"sse_f_02_14.sfs","iy/agency",	577.9793,	2376.8602,	2939.6063
"sse_f_02_16.sfs","iy/Gary",	538.9950,	2037.0008,	2349.4858
"sse_f_02_19.sfs","iy/tea",	572.1907,	2190.4421,	2744.6243
Word medial			
"sse_f_02_05.sfs","iy/people",	533.8898,	2297.5439,	2762.3784
"sse_f_02_06.sfs","iy/alleviate",	412.0224,	2108.4677,	2812.5599
"sse_f_02_07.sfs","iy/evening",	351.8525,	2425.5254,	2915.9789
"sse_f_02_10.sfs","iy/diseases",	420.0742,	2093.2929,	2707.7623
"sse_f_02_15.sfs","iy/field",	489.4434,	2162.2629,	2611.0818
"sse_f_02_17.sfs","iy/unbeatable",	455.4935,	2136.7220,	2748.7320
"sse_f_02_18.sfs","iy/leaves",	609.1624,	2063.6738,	2640.9705

A reasonable question to ask is whether there is a systematic difference in the formant frequencies of /i:/ across these two contexts (certainly F1 looks a bit higher and F2 looks a bit lower in word final position). For these data, the question we are asking is how likely it is that these two samples would have their means if they were really just two samples of the same underlying population of vowels. Thus the null hypothesis is that the observed variation in sample means is due to chance. If it turns out that the difference in means is unlikely just to be due to chance, then we can say that it is likely that there is a real effect.

To obtain the likelihood that a difference in sample means arose by chance we can simulate drawing samples of appropriate size from a single population and find out what proportion have a difference in means as large as the difference we observe in our data. This is just the calculation that is performed by the t-test.

We could perform a t-test on these data using a statistics package, but here we will just use the Excel spreadsheet program (the OpenOffice Calc spreadsheet has the same functions). You may need to install the optional "Analysis ToolPak" (sic) to get the statistical functions.

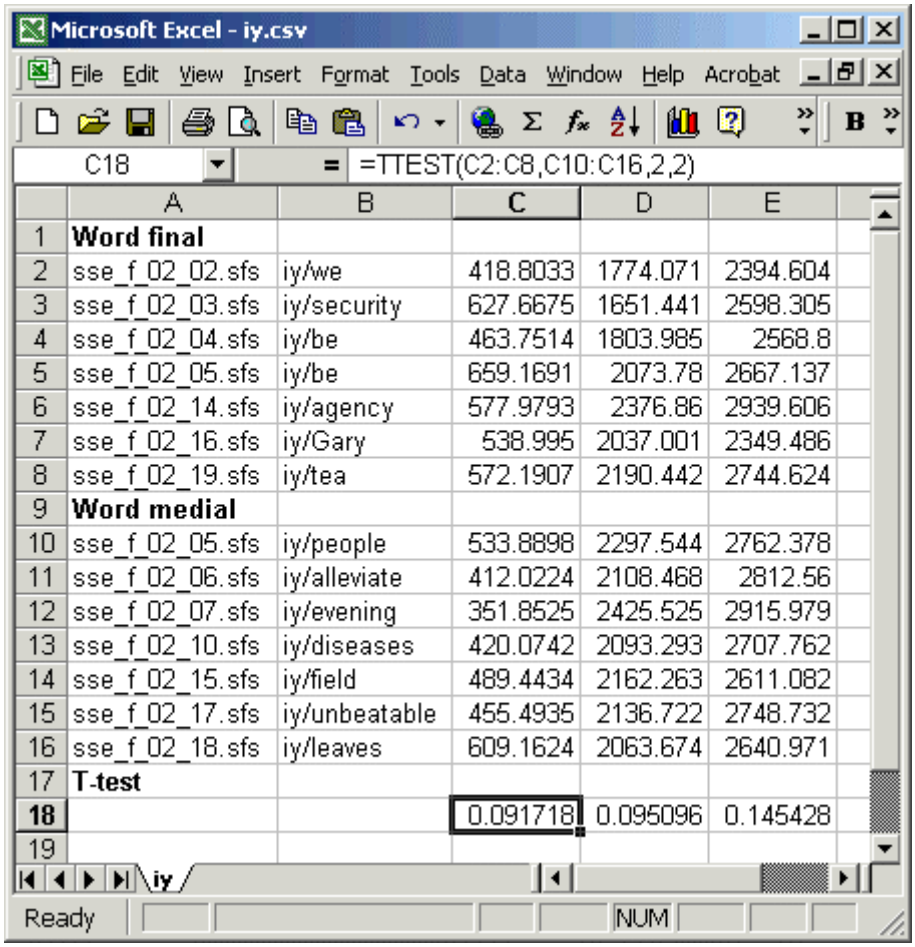
Figure D.5.1 shows the data in Excel, ready for the t-test values to be calculated:



	A	B	C	D	E
1	**Word final**				
2	sse_f_02_02.sfs	iy/we	418.8033	1774.071	2394.604
3	sse_f_02_03.sfs	iy/security	627.6675	1651.441	2598.305
4	sse_f_02_04.sfs	iy/be	463.7514	1803.985	2568.8
5	sse_f_02_05.sfs	iy/be	659.1691	2073.78	2667.137
6	sse_f_02_14.sfs	iy/agency	577.9793	2376.86	2939.606
7	sse_f_02_16.sfs	iy/Gary	538.995	2037.001	2349.486
8	sse_f_02_19.sfs	iy/tea	572.1907	2190.442	2744.624
9	**Word medial**				
10	sse_f_02_05.sfs	iy/people	533.8898	2297.544	2762.378
11	sse_f_02_06.sfs	iy/alleviate	412.0224	2108.468	2812.56
12	sse_f_02_07.sfs	iy/evening	351.8525	2425.525	2915.979
13	sse_f_02_10.sfs	iy/diseases	420.0742	2093.293	2707.762
14	sse_f_02_15.sfs	iy/field	489.4434	2162.263	2611.082
15	sse_f_02_17.sfs	iy/unbeatable	455.4935	2136.722	2748.732
16	sse_f_02_18.sfs	iy/leaves	609.1624	2063.674	2640.971
17	**T-test**				
18					
19					

Figure D.5.1 - Ready for t-test calculation in Excel

To enter the calculation for a t-test, use the TTEST(array1,array2,tails,type) function: array1 is the column of F1 values for word final, array2 is the column of F1 values for word medial, tails is 2 for a test in which we don't know whether the frequencies should be higher or lower, type is 2 for when we believe that both sets have the same variance. Figure D.5.2 shows the data in Excel, with the formula entered and copied under the F2 and F3 columns.



	A	B	C	D	E
1	**Word final**				
2	sse_f_02_02.sfs	iy/we	418.8033	1774.071	2394.604
3	sse_f_02_03.sfs	iy/security	627.6675	1651.441	2598.305
4	sse_f_02_04.sfs	iy/be	463.7514	1803.985	2568.8
5	sse_f_02_05.sfs	iy/be	659.1691	2073.78	2667.137
6	sse_f_02_14.sfs	iy/agency	577.9793	2376.86	2939.606
7	sse_f_02_16.sfs	iy/Gary	538.995	2037.001	2349.486
8	sse_f_02_19.sfs	iy/tea	572.1907	2190.442	2744.624
9	**Word medial**				
10	sse_f_02_05.sfs	iy/people	533.8898	2297.544	2762.378
11	sse_f_02_06.sfs	iy/alleviate	412.0224	2108.468	2812.56
12	sse_f_02_07.sfs	iy/evening	351.8525	2425.525	2915.979
13	sse_f_02_10.sfs	iy/diseases	420.0742	2093.293	2707.762
14	sse_f_02_15.sfs	iy/field	489.4434	2162.263	2611.082
15	sse_f_02_17.sfs	iy/unbeatable	455.4935	2136.722	2748.732
16	sse_f_02_18.sfs	iy/leaves	609.1624	2063.674	2640.971
17	**T-test**				
18			0.091718	0.095096	0.145428
19					

Figure D.5.2 - t-test calculation in Excel

What is the interpretation of the t-test probabilities? The t-test shows us that for each formant there is about a 1 in 10 chance (p about 0.1) that the difference in the sample means could have arisen even if they were actually from the same underlying population. This is not good evidence that there is a real effect here: we would expect the difference in observed means once in every ten experiments even if there were no effect to measure.

In this demonstration there was no obvious connection between the vowel contexts in each of the two sets, they were just two random samples of words in the recording. If we had planned the recording more carefully we could have constructed *paired contexts*, so that one of the pair would give us values in the first set, and one would give us a value in the second. For example we might have words such as "happy/happiness", "silly/sillyness" where we could compare the vowel formant frequencies with and without an additional affix. This is called a "paired" test and is fundamentally more sensitive than the independent samples test we reported above. In a paired test you are only looking for a systematic difference between members of each pair rather than a difference of means across the sets. In effect you are looking at the distribution of the hertz difference between the members of the pair, and the t-test establishes whether the mean of that difference across all pairs is significantly different from zero.

Comparison of means (2D)

A problem with the analysis of the last section is that we analysed separately any difference in the means of F1 and F2 and F3. If you think about it you can see that greater the number of parameters we check the more likely it is that we will find a low-probability random fluctuation. In other words we cannot just use a probability of 0.05 (say) to identify a significant event if we then apply the same significance separately to multiple parameters. The use of multiple parameters obliges us to look for a greater level of significance.

Another problem with treating F1 and F2 separately is that there may be a real effect which makes a small difference to **both** F1 and F2, but which is not significant when we test either separately.

Generally we need to consider a method to compare F1 and F2 together. First we'll look at graphing two samples on the F1-F2 plane, plotting a contour expressing one standard deviation in two-dimensions.

The following script collects F1 and F2 frequencies from the five long monophthongs from a single speaker. It then plots these on a scatter graph and estimates the shape and size of an ellipse that characterises the distribution of values.

```
/* f1f2distributions.sml - collect and plot F1-F2 distributions */

/* raw data tables - one per vowel type */
var        t1f1[1:1000],t1f2[1:1000],t1cnt;
var        t2f1[1:1000],t2f2[1:1000],t2cnt;
var        t3f1[1:1000],t3f2[1:1000],t3cnt;
var        t4f1[1:1000],t4f2[1:1000],t4cnt;
var        t5f1[1:1000],t5f2[1:1000],t5cnt;

/* output file */
file    gop;

/* calculate trimmed mean */
function var trimmean(table,len)
{
    var table[];    /* array of values */
    var len;        /* # values */
    var    i,j,tmp;
    var    lo,hi;

    /* sort table */
    for (i=2;i<=len;i=i+1) {
        j = i;
        tmp = table[j];
        while (table[j-1] > tmp) {
            table[j] = table[j-1];
            j = j - 1;
            if (j==1) break;
        }
        table[j] = tmp;
    }

    /* find mean over middle portion */
    lo = trunc(0.5 + 1 + len/5);    /* lose bottom 20% */
    hi = trunc(0.5 + len - len/5);    /* lose top 20% */
    j=0;
    tmp=0;
    for (i=lo;i<=hi;i=i+1) {
        tmp = tmp + table[i];
        j = j + 1;
    }
    if (j > 0) return(tmp/j) else return(ERROR);
}

/* get trimmed mean formant value for a segment */
function var measure_trimmed_mean(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */
```

```
    /* calculate trimmed mean */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(trimmean(af,nf));
}

/* record details of a single segment */
function var recordsegment(stime,etime,tf1,tf2,tcnt)
var    tf1[];
var    tf2[];
var tcnt;
{
    var stime,etime;
    var vf1,vf2

    vf1 = measure_trimmed_mean(stime,etime,5);
    vf2 = measure_trimmed_mean(stime,etime,8);
    if (vf1 && vf2) {
        tcnt=tcnt+1;
        tf1[tcnt] = vf1;
        tf2[tcnt] = vf2;
    }
}

/* plot ellipse for 2D gaussian */
function var plotellipse(sx,sy,sxy)
{
    stat sx,sy,sxy;
    var    xdata[0:360];
    var    ydata[0:360];
    var    a,cnt;
    var    rho;        /* correlation coeff */
    var    lam1,lam2;    /* eigenvalues */
    var    ax1,ax2;    /* axis lengths */
    var    alpha;        /* ellipse angle */

    /* get parameters of ellipse from distribution */
    rho = (sxy.mean-sx.mean*sy.mean)/(sx.stddev * sy.stddev);
    lam1 = 0.5 * (sx.variance+sy.variance+ \
        sqrt((sx.variance-sy.variance)*(sx.variance-sy.variance)+ \
        4*sx.variance*sy.variance*rho*rho));
    lam2 = 0.5 * (sx.variance+sy.variance- \
        sqrt((sx.variance-sy.variance)*(sx.variance-sy.variance)+ \
        4*sx.variance*sy.variance*rho*rho));
    ax1 = 2*sqrt(0.35*lam1);    /* about 70% inside */
    ax2 = 2*sqrt(0.35*lam2);
    alpha = 0.5*atan2(2*rho*sx.stddev*sy.stddev,sx.variance-
sy.variance);

    /* calculate locus of ellipse */
    cnt=0;
    for (a=0;a<=6.3;a=a+0.05) {
        xdata[cnt] = sx.mean + ax1*cos(a)*cos(alpha) -
ax2*sin(a)*sin(alpha);
        ydata[cnt] = sy.mean + ax1*cos(a)*sin(alpha) +
```

```
ax2*sin(a)*cos(alpha);
        cnt=cnt+1;
    }

    /* plot ellipse */
    plotparam("type=line");
    plotparam("char=");
    plotxdata(xdata,0);
    plot(gop,1,ydata,cnt);
}

/* plot F1-F2 for segment */
function var plotf1f2segment(label,tf1,tf2,tcnt)
var tf1[];
var tf2[];
{
    var        tcnt;
    string    label;
    stat    sx,sy,sxy;
    var        i;

    plotparam("type=point");
    plotparam("char="++label);

    /* plot raw samples */
    plotxdata(tf1,0);
    plot(gop,1,tf2,tcnt);

    /* collect statistics on samples */
    for (i=1;i<=tcnt;i=i+1) {
        sx += tf1[i];
        sy += tf2[i];
        sxy += tf1[i]*tf2[i];
    }

    /* plot ellipse at 1 standard deviation */
    plotellipse(sx,sy,sxy);
}

/* plot F1-F2 graph */
function var plotf1f2()
{
    openout(gop,"|dig -g -s 500x375 -o c:/tmp/dig.gif");
    /* openout(gop,"|dig"); */
    plottitle(gop,"Vowel variability");
    plotparam("title=single speaker, all contexts");
    plotparam("xtitle=F1 Frequency (Hz)");
    plotparam("ytitle=F2 Frequency (Hz)");
    plotparam("type=point");
    plotaxes(gop,1,200,800,500,2500);

    /* for each segment in turn */
    plotf1f2segment("i",t1f1,t1f2,t1cnt);
    plotf1f2segment("u",t2f1,t2f2,t2cnt);
    plotf1f2segment("3",t3f1,t3f2,t3cnt);
    plotf1f2segment("A",t4f1,t4f2,t4cnt);
    plotf1f2segment("O",t5f1,t5f2,t5cnt);

    close(gop);
}
```

```
/* for each file to be processed */
main {
    var    num;    /* # annotated regions */
    var    i;
    var stime,etime;
    var    scode;
    string    label;

    /* report file */
    num = numberof(".");
    print#stderr "File ",$filename," has ",num," annotations\n";

    /* for each annotation */
    for (i=1;i<=num;i=i+1) {
        stime = timen(".",i);
        etime = stime + lengthn(".",i);
        label = matchn(".",I);
        switch (label) {
        case "i:": recordsegment(stime,etime,t1f1,t1f2,t1cnt);
        case "u:": recordsegment(stime,etime,t2f1,t2f2,t2cnt);
        case "3:": recordsegment(stime,etime,t3f1,t3f2,t3cnt);
        case "A:": recordsegment(stime,etime,t4f1,t4f2,t4cnt);
        case "O:": recordsegment(stime,etime,t5f1,t5f2,t5cnt);
        }
    }
}

/* summarise collected data */
summary {
    /* plot graph */
    plotf1f2();
}
```

Figure D.5.3 shows the output of the script. These data seem rather noisy.

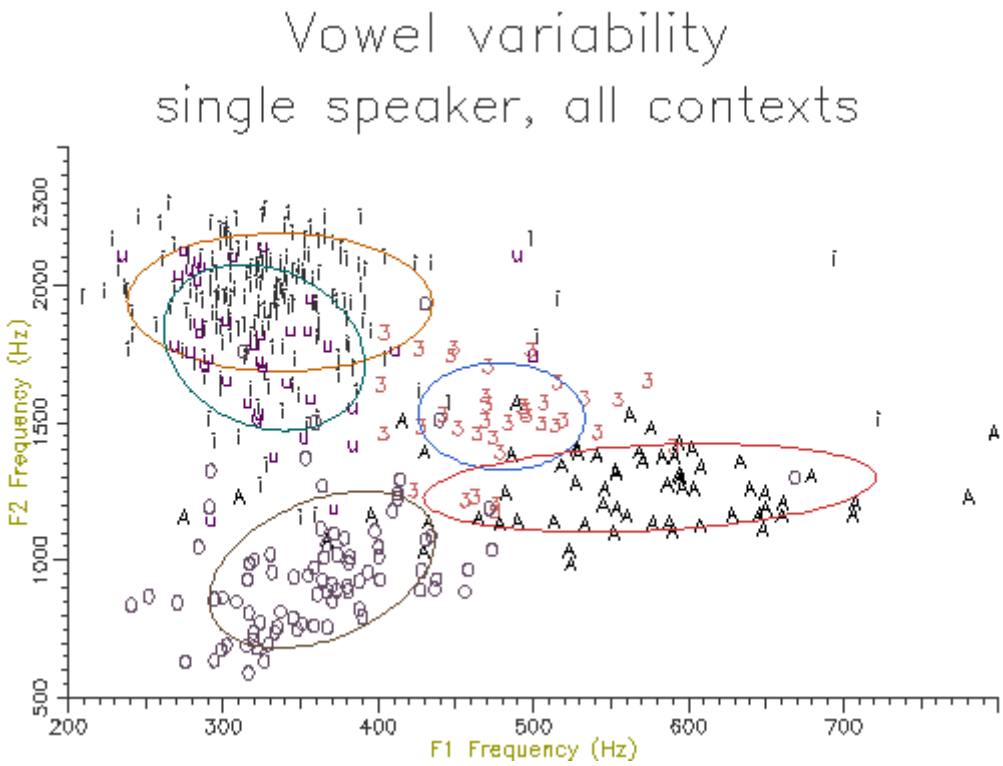


Figure D.5.3 - Vowel distribution on F1-F2 plane

Comparing the centroids of two vowel samples on the F1-F2 plane

Looking carefully at Figure D.5.3 you will see that there is a large overlap in the distributions of /i:/ and /u:/ on the F1-F2 plane. We will now consider how we might perform a statistical test that determines whether these two samples are really distinct or come from the same underlying population (we are pretty sure that /i:/ and /u:/ are different, of course!). Rather than use two t-tests on the F1 values and the F2 values separately, we will show how to perform a single test that uses F1 and F2 *jointly*.

A statistical test on one random variable is called a "univariate" test, while a statistical test on 2 or more variables is called "multivariate". In this case we need to use the multivariate equivalent of the t-test, and this is called "Hotelling's T^2 test".

The script below collects information about the F1 and F2 distribution of two vowels and calculates the value of Hotelling's T^2 statistic on the two samples.

```
/* f1f2compare.sml - calculate Hotelling's T-squared on vowel pair */

/* vowels to measure */
string    label1;
string    label2;

/* raw data tables - one per vowel type */
var        t1f1[1:1000],t1f2[1:1000],t1cnt;
var        t2f1[1:1000],t2f2[1:1000],t2cnt;

/* calculate trimmed mean */
function var trimmean(table,len)
{
    var table[];    /* array of values */
    var len;        /* # values */
    var    i,j,tmp;
    var    lo,hi;

    /* sort table */
    for (i=2;i<=len;i=i+1) {
        j = i;
        tmp = table[j];
        while (table[j-1] > tmp) {
            table[j] = table[j-1];
            j = j - 1;
            if (j==1) break;
        }
        table[j] = tmp;
    }

    /* find mean over middle portion */
    lo = trunc(0.5 + 1 + len/5);    /* lose bottom 20% */
    hi = trunc(0.5 + len - len/5);    /* lose top 20% */
    j=0;
    tmp=0;
    for (i=lo;i<=hi;i=i+1) {
        tmp = tmp + table[i];
        j = j + 1;
    }
    if (j > 0) return(tmp/j) else return(ERROR);
}

/* get trimmed mean formant value for a segment */
function var measure_trimmed_mean(stime,etime,pno)
{
    var    stime;    /* start time */
    var    etime;    /* end time */
```

```
    var    pno;    /* FM parameter # */
    var    t;        /* time */
    var af[1:1000];    /* array of values */
    var    nf;        /* # values */

    /* calculate trimmed mean */
    nf=0;
    t=next(FM,stime);
    while ((t < etime)&&(nf < 1000)) {
        nf = nf+1;
        af[nf] = fm(pno,t);
        t = next(FM,t);
    }

    return(trimmean(af,nf));
}

/* record details of a single segment */
function var recordsegment(stime,etime,tf1,tf2,tcnt)
var    tf1[];
var    tf2[];
var tcnt;
{
    var stime,etime;
    var vf1,vf2

    vf1 = measure_trimmed_mean(stime,etime,5);
    vf2 = measure_trimmed_mean(stime,etime,8);
    if (vf1 && vf2) {
        tcnt=tcnt+1;
        tf1[tcnt] = vf1;
        tf2[tcnt] = vf2;
    }
}

/* calculate mean and covariance */
function var covar(a1,a2,cnt,amean,acov)
var amean[],acov[];
{
    var        a1[],a2[],cnt;
    var     i;
    stat    s1,s2;

    for (i=1;i<=cnt;i=i+1) {
        s1 += a1[i];
        s2 += a2[i];
    }
    amean[1]=s1.mean;
    amean[2]=s2.mean;
    for (i=1;i<=cnt;i=i+1) {
        a1[i] = a1[i] - s1.mean;
        a2[i] = a2[i] - s2.mean;
    }
    clear(acov);
    for (i=1;i<=cnt;i=i+1) {
        acov[1] = acov[1] + a1[i]*a1[i];
        acov[2] = acov[2] + a1[i]*a2[i];
        acov[3] = acov[3] + a2[i]*a1[i];
        acov[4] = acov[4] + a2[i]*a2[i];
    }
    for (i=1;i<=4;i=i+1) acov[i] = acov[i] / (cnt-1);
```

```
}

/* calculate Hotelling's T-squared */
function var hotelling(v11,v12,v1cnt,v21,v22,v2cnt)
var    v11[],v12[],v1cnt;
var    v21[],v22[],v2cnt;
{
    var    i;
    var mean1[1:2];
    var    cov1[1:4];
    var mean2[1:2];
    var cov2[1:4];
    var    dmean[1:2];
    var    cov[1:4];
    var    icov[1:4];
    var    t2;

    /* get individual means and covariances */
    covar(v11,v12,v1cnt,mean1,cov1);
    covar(v21,v22,v2cnt,mean2,cov2);

    /* get difference in means */
    dmean[1] = mean2[1]-mean1[1];
    dmean[2] = mean2[2]-mean1[2];

    /* combine covariances */
    for (i=1;i<=4;i=i+1) {
        cov[i] = ((v1cnt-1)*cov1[i] + (v2cnt-1)*cov2[i])/(v1cnt+v2cnt-2);
    }

    /* invert covariance */
    icov[1]=cov[4]/(cov[4]*cov[1]-cov[2]*cov[3]);
    icov[2]=(1-cov[1]*icov[1])/cov[2];
    icov[3]=(1-cov[1]*icov[1])/cov[3];
    icov[4]=cov[1]*icov[1]/cov[4];

    /* put t2 together in parts */
    t2 = 0;
    t2 = t2 + dmean[1]*(dmean[1]*icov[1]+dmean[2]*icov[3]);
    t2 = t2 + dmean[2]*(dmean[1]*icov[2]+dmean[2]*icov[4]);
    t2 = (v1cnt*v2cnt*t2)/(v1cnt+v2cnt);

    return(t2);
}

/* get names of vowels to measure */
init {
    print "Enter vowel label 1 : ";
    input label1;
    print "Enter vowel label 2 : ";
    input label2;
}

/* for each file to be processed */
main {
    var    num;    /* # annotated regions */
    var    i;
    var stime,etime;
    var    scode;
    string    label;
```

```
    /* report file */
    num = numberof(".");
    print#stderr "File ",$filename," has ",num," annotations\n";

    /* for each annotation */
    for (i=1;i<=num;i=i+1) {
        stime = timen(".",i);
        etime = stime + lengthn(".",i);
        label = matchn(".",i);
        if (compare(label,label1)==0) {
            recordsegment(stime,etime,t1f1,t1f2,t1cnt);
        }
        else if (compare(label,label2)==0) {
            recordsegment(stime,etime,t2f1,t2f2,t2cnt);
        }
    }
}

/* summarise collected data */
summary {
    var t,f;

    print "Analysis summary:\n"
    print "  Number of files = ",$filecount:1,"\n";
    print "  Number of instances of /",label1,"/ = ",t1cnt:1,"\n";
    print "  Number of instances of /",label2,"/ = ",t2cnt:1,"\n";

    /* calculate Hotelling T-squared statistic */
    t = hotelling(t1f1,t1f2,t1cnt,t2f1,t2f2,t2cnt);
    print "  Hotelling T-squared statistic =",t:5:2,"\n";

    /* report equivalent F statistic */
    f=(t1cnt+t2cnt-3)*t/((t1cnt+t2cnt-2)*2);
    print "  For significance find probability that F(2,", \
            (t1cnt+t2cnt-3):1,") >",f:5:2,"\n";
}
```

When this script is run on the data used to produce Figure D.5.3 on the vowels /i:/ and /u:/, the result is:

```
Analysis summary:
  Number of files = 30
  Number of instances of /i:/ = 192
  Number of instances of /u:/ = 39
  Hotelling T-squared statistic = 18.80
  For significance find probability that F(2,228) > 9.36
```

From this output you can see that the Hotelling's T^2 statistic is 18.8 for these two vowels. To interpret this number we need to know how often this value would occur for samples of the size we used if there were no underlying difference between these vowels. To turn the T^2 statistic into a probability we can use the fact that its distribution follows the same shape as the F distribution of a particular configuration (basically for a *p*-variate sample of df= *n*, the F statistic is (n-p)/p(n-1) times the T^2 statistic). So in this case, the likelihood of a T^2 of 18.8 is equivalent to the likelihood of an F(2,228) statistic of 9.36. Since in general we are only interested in the significance of the statistic, here are a few critical values from the F-distribution F(2,n):

F statistic	p < 0.05	p < 0.01
F(2,10)	4.1	7.56
F(2,20)	3.49	5.85
F(2,50)	3.19	5.08
F(2,100)	3.10	4.85
F(2,200)	3.06	4.77

F(2,500)	3.04	4.71

Since the likelihood of F(2,200) being greater than 4.77 is less than 0.01, we can be sure that the likelihood of F(2,228) being greater than 9.36 is much less than 0.01. Thus there is a significant difference between these two distributions (phew!).

The procedure above could be extended to work with F1, F2 and F3 if required. But if you wanted to test if 3 or more samples came from the same population (rather than just 2 in this case) you would need to perform a multivariate analysis of variance or MANOVA.

Bibliography

The following have been used in the development of this tutorial:

- Maria Isabel Ribeiro, Gaussian Probability Density Functions: Properties and Error Characterization at http://omni.isr.ist.utl.pt/~mir/pub/probability.pdf
- Hotelling's T^2 test at http://www.itl.nist.gov/div898/handbook/pmc/section5/pmc543.htm
- Critical F Distribution Calculator at http://www.psychstat.smsu.edu/introbook/fdist.htm

---

Appendix E HTK Hidden Markov modelling toolkit with SFS

This document provides a tutorial introduction to the use of SFS in combination with the Cambridge Hidden Markov modelling toolkit (HTK) for pattern processing of speech signals. The tutorial covers installation, file conversion, phone and phone-class recognition, phonetic alignment, pronunciation variation analysis and, of particular interest for current purposes, automatic dysfluency analysis. The work preceding automatic dysfluency analysis contains material that is essential for understanding this topic and could be used in developments of the recognizer (e.g. pronunciation variant analysis). Some background understanding of the Unix command line interpreter is assumed and basic knowledge of how Hidden Markov models work. This tutorial refers to versions 4.6 and later of SFS with version 3.3 of HTK.

1. Installation, Acquiring and Chunking the audio signal

These installation instructions refer to Windows computers. However most of the tutorial applies to other platforms where HTK and SFS command-line programs can be run under a Unix-like shell program.

Installation of CYGWIN

CYGWIN provides a Unix-like programming environment for Windows computers. This environment will be used in the tutorial so that scripts for processing multiple files using the BASH shell language can be used. This is useful because the shell language is simple yet powerful and runs on many different computing platforms.

CYGWIN can be downloaded from the CYGWIN home page at www.cygwin.com. From there download and run the program `setup.exe` which manages the installation of CYGWIN. This program first collects information about your nearest CYGWIN distribution site then presents you with a list of components to install. Finally it downloads and installs the selected components. The same program can be used to update your installation as new software versions become available and to add/delete components. The setup program goes through these steps:

1. **Choose installation type**. Choose "Install from Internet" if you have a reliable internet connection. Otherwise choose "Download from Internet" to copy the files onto your computer and then "Install from Local Directory" to install them.
2. **Choose installation directory**. We suggest you leave this as "C:/cygwin" unless you know what you are doing.
3. **Select local package directory**. Put directory (aka "folder") here, where CYGWIN will put the downloaded files before installation. You could enter a temporary directory name. We use "C:/download/cygwin". You may need to make the folders first using Windows Explorer.
4. **Select Internet connection**. Leave as default.
5. **Choose a download site**. Highlight an address in the list that seems to come from your own country. We choose "ftp://ftp.mirror.ac.uk/".
6. **Select packages**. Use this page to investigate what packages (components) are available for download. Many are rather old and obscure elements of the Unix operating system. For the purposes of this tutorial you should download at least the following:
    1. All components in the "Base" category.
    2. Devel|BINUTILS: The GNU assembler, linker and binary utilities
    3. Devel|GCC: C Compiler
    4. Devel|GCC-G++: GCC C++ compiler
    5. Devel|MAKE: the GNU version of the 'make' utility

    But feel free to install any of the other goodies that take your fancy.
7. **Download**. The program then downloads and installs the selected packages.
8. **Installation complete**. Choose both boxes to put a CYGWIN icon on your desktop and put a CYGWIN entry in the Start Menu.

After installation your should see a CYGWIN icon on the desktop and a Start menu option Start|Programs|Cygwin|Cygwin BASH shell. Either of these will start up a command window which provides a Unix-like environment in which we will be demonstrating SFS and HTK.

A good introduction to programming the Unix environment can be found in the old but essential "The Unix Programming Environment" by Kernighan and Pike.
Available at Amazon.co.uk.

It is worth exploring the CYGWIN environment to get used to the way it maps the names of the Windows disks and folders. A folder like "C:\WINDOWS" is referred to as "c:/windows" in CYGWIN, or as "/cygdrive/c/windows". Your home directory in CYGWIN (referred to as "~") will actually be a subdirectory of the windows folder c:\cygwin\home.

Installation of a Text Editor

You will need a suitable text editor for editing scripts and other text files in this tutorial. Our recommendation is **TextPad** which can be downloaded from www.textpad.com. This is a shareware program which requires registration if you use it extensively.

Installation of HTK

The Cambridge University Hidden Markov modelling toolkit (HTK) can be downloaded from htk.eng.cam.ac.uk. To check that you have read the licence conditions, they ask you first to register your name and e-mail address with them. They will then send you a password to use to download the HTK sources from htk.eng.cam.ac.uk/ftp. For the purposes of this tutorial we downloaded http://htk.eng.cam.ac.uk/ftp/beta/HTK-3.3-alpha1.tar.gz into our cygwin home directory. By the time you read this tutorial it is very likely that there will be a new release with a different filename. The CYGWIN command to unpack this is just:

```
$ tar xvfz HTK-3.3-alpha1.tar.gz
```

When unpacked, a sub-directory called "htk" will be created under your home directory. You can now delete the downloaded distribution file.

To build HTK for CYGWIN you first need to set a number of environment variables. We suggest you create a file called "htk.env" in your home directory containing the following:

```
export HTKCF='-O2 –DCYGWIN'
export HTKLF='-o a.out'
export HTKCC='gcc'
export HBIN='..'
export Arch=ASCII
export CPU=cygwin
export PATH=~/htk/bin.cygwin:$PATH
```

Then each time you want to use HTK you can just type

```
$ source htk.env
```

Alternatively you can put these commands in a file ".bash_login" in your home directory so that they will be executed each time you log in.

Unfortunately, as of the date of writing this tutorial, the HTK distribution needs patching before it can be compiled under CYGWIN. These are the follwing edits that you need to make using a text editor:

1. Edit HTKLib/HShell.h and include at the end of the file:

```
#ifdef CYGWIN
#include <asm/socket.h>
#endif
```

5. Edit HTKLib/HGraf.null.c and include at the end of the file:

```
/* EXPORT HTextHeight: return the height of s in pixels */
int HTextHeight(char *str)
{
  return 0;
}
```

11. Edit HTKTools/makefile and remove the reference to "-lX11" in the instructions for HSLab:

```
HSLab:  $(hlib)/HTKLib.$(CPU).a HSLab.o
```

13. $(CC) HSLab.o $(HLIBS) -lm $(HTKLF)
14. mv a.out $(HBIN)/bin.$(CPU)/HSLab

We can now make HTK with the following instructions:

```
$ source htk.env
$ cd ~/htk
$ mkdir bin.cygwin
$ cd HTKLib
$ cp HGraf.null.c HGraf.c
$ make
$ cd ../HTKTools
$ make
$ cd ../HLMLib
$ make
$ cd ../HLMTools
$ make
```

Installation of SFS

As mentioned earlier, SFS can be downloaded from www.phon.ucl.ac.uk/resource/sfs/. Run the installation package and select the option "Add SFS to command-line path" to add the SFS program directory to the search path for programs to run from the command prompt and the CYGWIN shell. You may need to reboot for this change to take effect.

2. Phone-class recognition

In this section we will describe a "warm-up" exercise to show how SFS and HTK can be used together to solve a simple problem. The idea is to demonstrate the software tools rather than to achieve ultimate performance on the task.

We will demonstrate the use of SFS and HTK to build a system that automatically labels an audio signal with annotations which divide the signal into regions of "silence", "voiced speech", and "voiceless speech".

Source data

For this demonstration we will use some annotated data that are part of the SCRIBE corpus (see www.phon.ucl.ac.uk/resource/scribe). These data are interesting because they contain some "acoustic" level annotations - that is phonetic annotation at a finer level of detail than normal. In particular the annotation marks voiced and voiceless regions *within* phonetic segments. An example is shown in Figure E.2.1.

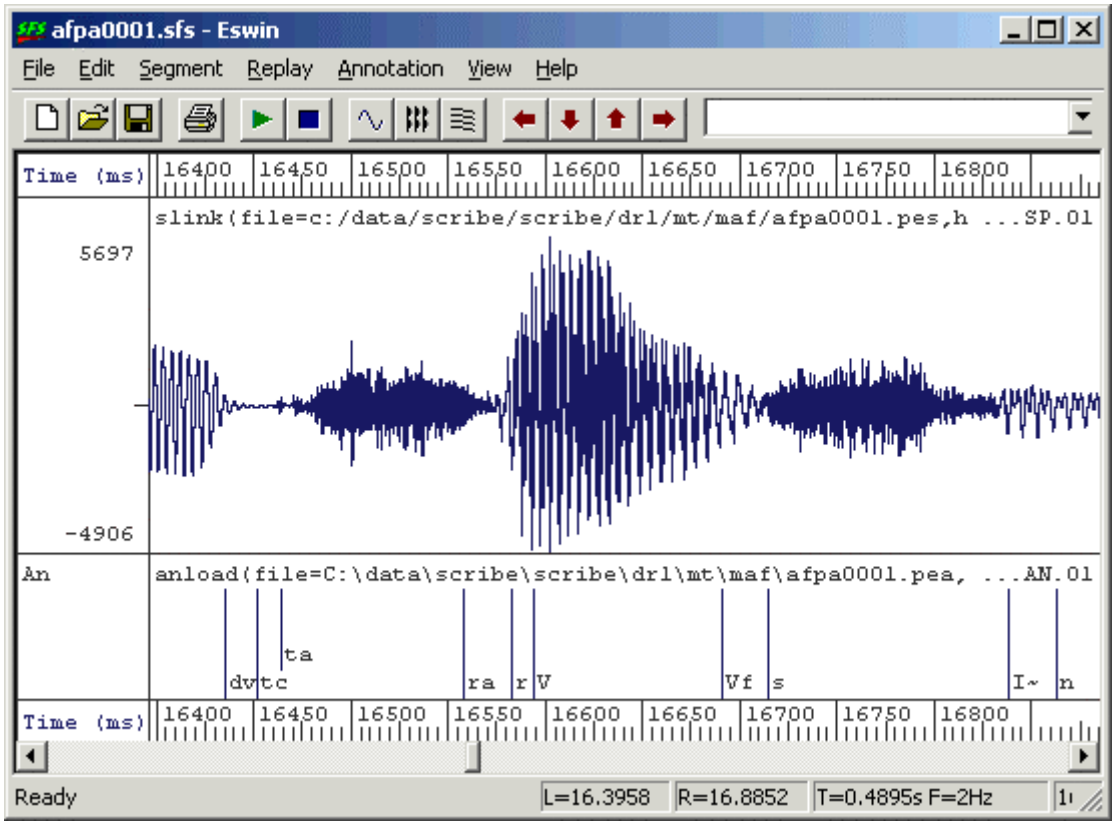


Figure E.2.1 - Acoustic level annotations

The files we will be using are from the 'many talker' sub-corpus and are as follows

Speaker	Training Signal	Label	Testing Signal	Label
mac	Acpa0002.pes	Acpa0002.pea	acpa0001.pes	acpa0001.pea
	acpa0003.pes	acpa0003.pea		
	acpa0004.pes	acpa0004.pea		
mae	Aepa0001.pes	Aepa0001.pea	aepa0002.pes	aepa0002.pea
	aepa0003.pes	aepa0003.pea		
	aepa0004.pes	aepa0004.pea		
maf	Afpa0001.pes	Afpa0001.pea	afpa0003.pes	afpa0003.pea
	afpa0002.pes	afpa0002.pea		
	afpa0004.pes	afpa0004.pea		
mah	Ahpa0001.pes	Ahpa0001.pea	ahpa0004.pes	ahpa0004.pea
	ahpa0002.pes	ahpa0002.pea		
	ahpa0003.pes	ahpa0003.pea		
mam			ampa0001.pes	ampa0001.pea
			ampa0002.pes	ampa0002.pea
			ampa0003.pes	ampa0003.pea
			ampa0004.pes	ampa0004.pea

You can see from this table that we have reserved one quarter of each training speaker's recording for testing, and one whole unseen speaker. This means that we can test our parser on material that has not been used for training and also on a speaker that has not been used for training. This should give us a more robust estimate of the recognizer's performance.

Loading source data into SFS

We will start by setting up SFS files which point to the SCRIBE data. We will make a new directory in our cygwin directory and run a script which makes the SFS files. The SCRIBE audio files are in a raw binary format at a sampling rate of 20,000 samples/sec. We "link" these into the SFS files rather than waste disk space by copying them. The SCRIBE label files are in SAM format, which the SFS program anload can read (with "-S" switch). The shell script is as follows:

```
# doloadsfs.sh - load scribe data into new SFS files
```

```
for s in c e f h m
do
  for f in 0001 0002 0003 0004
  do
    hed -n ma$s.$f.sfs
    slink -isp -f 20000 c:/data/scribe/scribe/dr1/mt/ma$s/a${s}pa$f.pes \
                                        ma$s.$f.sfs
    anload -S c:/data/scribe/scribe/dr1/mt/ma$s/a${s}pa$f.pea ma$s.$f.sfs
  done
done
```

We'd run this in its own subdirectory as follows:

```
$ mkdir tutorial1
$ cd tutorial1
$ sh doloadsfs.sh
```

Data Preparation

There are two data preparation tasks: designing and computing a suitable acoustic feature representation of the audio files so they are suitable for the recognition task and mapping the annotation labels into a suitable set of symbols.

For the first task, a simple spectral envelope feature set would seem to be adequate. We will try this first and develop alternatives later. The SFS program voc19 performs a 19-channel filterbank analysis on an audio signal. It consists of 19 band-pass filters spaced on a bark scale from 100 to 4000Hz. The outputs of the filters are rectified, low-passed filtered at 50Hz, resampled at 100 frames/second and finally log-scaled. To run voc19 on all our training and testing data we type:

```
$ apply voc19 ma*.sfs
```

For the second task, we are aiming to label the signal with three different labels, according to whether there is silence, voiced speech or voiceless speech. Let us lable these three types as SIL, VOI, UNV. Our annotation preparation task is to map existing annotations to these types. In this case we are not even sure of the inventory of symbols used by the SCRIBE labelers, so we write a script to collect the names of all the different annotations they used:

```
/* ancollect.sml -- collect inventory of labels used */

/* table to hold annotation labels */
string    table[1:1000];
var    tcount;

/* function to check/add label */
function var checklabel(str)
{
   string str;

   if (entry(str,table)) return(0);
   tcount=tcount+1;
   table[tcount]=str;
   return(1);
}

/* for each input file */
main {
   var    i,num;

   num=numberof(".");
   for (i=1;i<=num;i=i+1) checklabel(matchn(".",i));
}

/* output sorted list */
summary {
   var    i,j;
   string    t;
```

```
    /* insertion sort */
    for (i=2;i<=tcount;i=i+1) {
        j=i;
        t=table[j];
        while (compare(t,table[j-1])<0) {
            table[j] = table[j-1];
            j=j-1;
            if (j==1) break;
        }
        table[j]=t;
    }

    /* output list */
    for (i=1;i<=tcount;i=i+1) print table[i],"\n";
}
```

To run this script from the CYGWIN shell, we type:

```
$ sml ancollect.sml ma*.sfs >svumap.txt
```

We now need to edit the file svumap.txt so as to assign each input annotation with a new SIL, VOI or UNV annotation. Here is the start of the file after editing:

```
# SIL
## SIL
%tc SIL
+ SIL
/ SIL
3: VOI
3:? UNV
3:a UNV
3:af UNV
3:f UNV
3:~ VOI
=l VOI
=lx VOI
=lx? VOI
=lxf UNV
=lxf0 UNV
=m VOI
=mf UNV
=n VOI
...
```

We now translate the SCRIBE labels to the new 3-way classification. We use the SFS anmap program with the "-m" option to collapse adjacent repeated symbols into one instance:

```
$ apply "anmap -m svumap.txt" ma*.sfs
```

The result of the mapping can be plainly seen in Figure E.2.2.

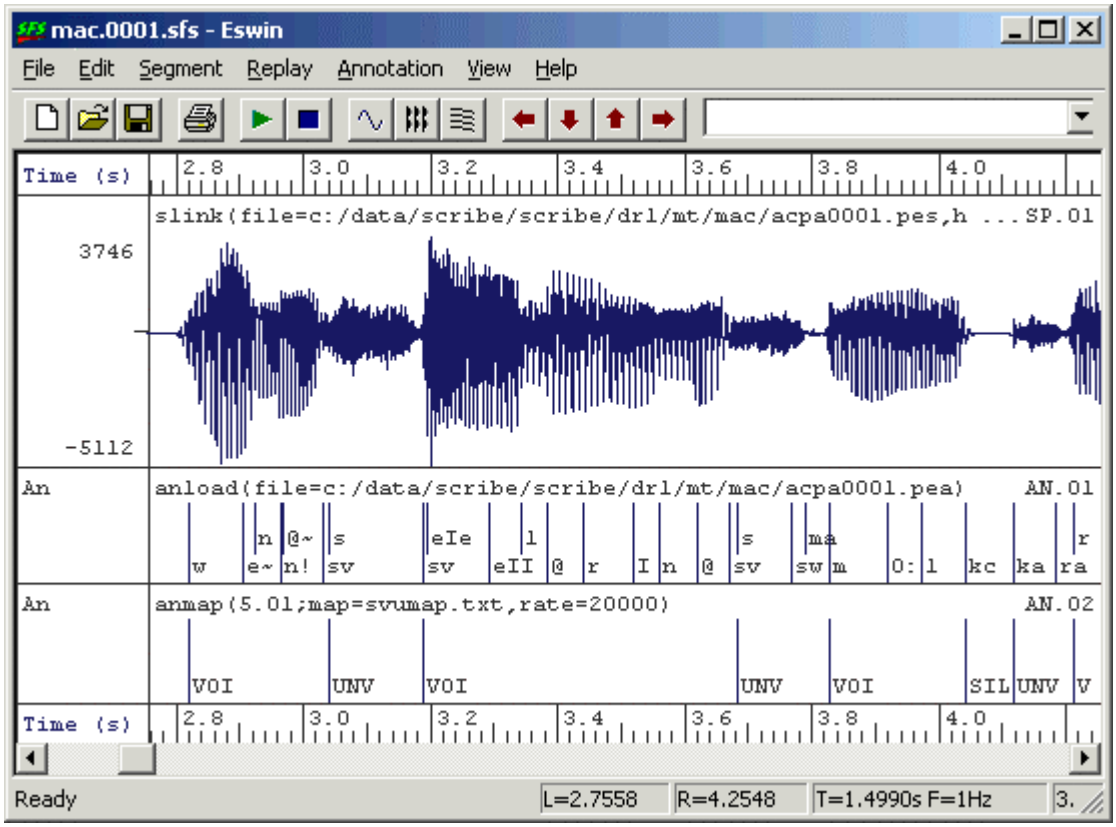


Figure E.2.2 - Mapped annotations

Export of data to HTK format

Unfortunately, HTK cannot (as yet) read SFS files directly, so the next step is to export the data into files formatted so that HTK can read them. Fortunately, SFS knows how to read and write HTK formatted files. To write a set of HTK data files from the voc19 analysis we performed, just type:

$ apply "colist -H" ma*.sfs

This creates a set of HTK format data files with names modelled after the SFS files. Similarly to write a set of HTK format label files, use:

$ apply "anlist -h -O" ma*.sfs

We now have a set of data files ma*.dat and a set of label files ma*.lab in HTK format ready to train some HMMs.

HTK configuration

For training HMMs, HTK requires us to build some configuration files beforehand. The first file is a general configuration of all HTK tools. We will put this into a file called config.txt

```
# config.txt - HTK basic parameters
SOURCEFORMAT = HTK
TARGETKIND = FBANK
NATURALREADORDER = T
```

In this file we specify that the source files are in HTK format, that the training data are already processed into filterbank parameters, and that the data are stored in the natural byte order for the machine.

The second configuration file we need is a prototype hidden Markov model which we will use to create the models for the three different labels. This configuration file is specific to a one-state HMM with a 19-dimensional observation vector, so we save it in a file called proto-1-19.hmm

```
<BeginHMM>
<NumStates> 3 <VecSize> 19 <FBANK>
 <State> 2
  <Mean> 19
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
```

```
 <Variance> 19
  1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <TransP> 3
 0.0 1.0 0.0
 0.0 0.9 0.1
 0.0 0.0 1.0
<EndHMM>
```

It is confusing that an HTK configuration file for a 1-state HMM has 3 states, but the HTK convention is that the first state and the last state are *non-emitting*, that is they are part of the network of HMMs but do not describe any of the input data. Only the middle state of the three emits observation vectors. Briefly this HMM definition file sets up a model with a single state which holds the mean and variance of the 19 filterbank channel parameters. The transition probability matrix simply says that state 1 always jumps to state 2, that state 2 jumps to state 3 with a probability of 1 in 10, and that state 3 always jumps to itself.

HTK training

To train the HMMs we need a file containing a list of all the data files we will use for training. Here we call it train.lst, and it has these contents:

```
mac.0002.dat
mac.0003.dat
mac.0004.dat
mae.0001.dat
mae.0003.dat
mae.0004.dat
maf.0001.dat
maf.0002.dat
maf.0004.dat
mah.0001.dat
mah.0002.dat
mah.0003.dat
```

We can now use the following script to train the HMMs:

```
# dotrain.sh
for s in SIL VOI UNV
do
   cp proto-1-19.hmm $s.hmm
   HRest -T 1 -C config.txt -S train.lst -l $s $s.hmm
Done
```

In this script we copy our prototype HMM into SIL.hmm, VOI.hmm, UNV.hmm and then train these HMMs individually using the portions of the data files that are labeled with SIL, VOI and UNV respectively.

HTK testing

To evaluate how well our HMMs can label an unseen speech signal with the three labels, we recognise the data we reserved for testing and compare the recognised labels with the labels we generated. We will show how to do the recognition in this section, and look at performance evaluation in the next.

To recognise a data file using our trained HMMs we need three further HTK configuration files and a list of test files. The first configuration file we need is just a list of the HMMs; store this in a file called phone.lst:

```
SIL
VOI
UNV
```

The next file we need is a dictionary that maps word pronunciations to a sequence of phone pronunciations. This sounds rather odd in this application, but is necessary because HTK is set up to recognise words rather than phones. Since we are really building a kind of phone recogniser, we solve the problem by having a dictionary of exactly three "words" each of which is "pronounced" by one of the phone labels. Put this into a file called phone.dic:

```
SIL    SIL
VOI    VOI
```

```
UNV    UNV
```

The last configuration file we need is a recognition grammar that describes which HMM sequences are allowed and what the probability is that each model should follow one of the others. For now, we will just use a default grammar consisting of a simple "phone loop" where the symbols can come in any order and with equal likelihood. This kind of grammar can be built using the HTK program HBuild from the phone list, as follows:

$ HBuild phone.lst phone.net

The resulting grammar file phone.net looks like this:

```
VERSION=1.0
N=7    L=9
I=0    W=!NULL
I=1    W=!NULL
I=2    W=SIL
I=3    W=VOI
I=4    W=UNV
I=5    W=!NULL
I=6    W=!NULL
J=0    S=0    E=1    l=0.00
J=1    S=5    E=1    l=0.00
J=2    S=1    E=2    l=-1.10
J=3    S=1    E=3    l=-1.10
J=4    S=1    E=4    l=-1.10
J=5    S=2    E=5    l=0.00
J=6    S=3    E=5    l=0.00
J=7    S=4    E=5    l=0.00
J=8    S=5    E=6    l=0.00
```

Do not worry too much about this file. It looks more complex than it really is - basically the lines starting with "I" represent nodes of a simple transition network, while the lines starting with "J" represent arcs that run from one node to another and have a transition probability (stored as a log likelihood).

The last thing we need is a list of test filenames; put this in train.lst:

```
mac.0001.dat
mae.0002.dat
maf.0003.dat
mah.0004.dat
mam.0001.dat
mam.0002.dat
mam.0003.dat
mam.0004.dat
```

Recognition can be performed with the following script. We run HVite to generate a set of recognised label files (with a .rec file extension) then load them into the SFS files for evaluation:

```
# dotest.sh
Hvite -T 1 -C config.txt -w phone.net -o S -S test.lst phone.dic phone.lst
for f in `cat test.lst`
do
   g=`echo $f | sed s/.dat//`
   anload -h $g.rec $g.sfs
done
```

In this script, the HVite option "-w phone.net" instructs it to perform word recognition, while the option "-o S" requests it not to put numeric scores in the recognised label files. The loop at the bottom takes the name of each test data file in turn, strips off the .dat from its name and loads the .rec file into the .sfs file with the SFS program anload.

Performance evaluation

We are now in a position to evaluate how well our system is able to divide a speech signal up into SIL-VOI-UNV. The inputs to the evaluation process are the SFS files for the test data. These now have the original acoustic annotations, the mapped annotations and the recognised annotations, see Figure E.2.3.

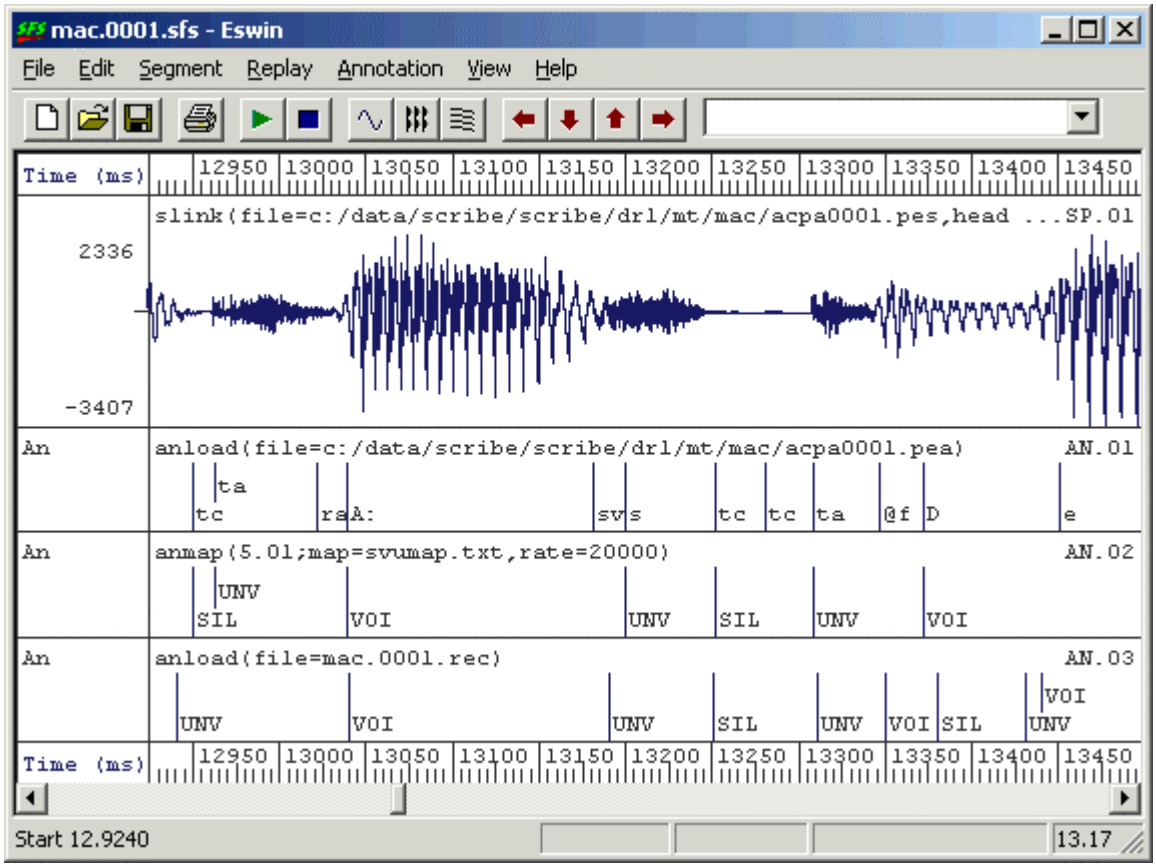


Figure E.2.3 - Recognition results

The figure shows that the recognised annotations are not bad, but there are some mistakes. We now need to consider what numerical measure we might use to describe the performance of the recogniser. In this case a suitable measure is the *frame labeling rate* (the percentage of frames of the signal which have been classified correctly). Since the recognised labels have been based on a frame rate of 100 frames per second, it also makes sense to evaluate the recognition performance at this rate.

The SFS program ancomp compares two annotation sets in various ways. It is supplied with a reference set and a test set and it can compare the timing of the labels, the content of the labels or the sequence of the labels. Here we want to look at the content of the labels every 10ms and build a confusion matrix that identifies which input (=reference) labels have been mapped to which output (=test) labels with what frequency. The command and some output for a single test file follows:

```
$ ancomp -r an.02 -t an.03 -f mac.0001.sfs
```

	SIL	UNV	VOI
SIL:	1114	31	6
UNV:	57	639	100
VOI:	140	138	1514

The "-f" switch to ancomp selects the frame labeling mode of comparison, while the switches "-r" and "-t" specify the reference and the test annotation items respectively. This way of running ancomp delivers performance on a single file only however, and we would like to get the performance over all test files. We can do this by asking ancomp to dump its raw label comparisons to its output and then combine the outputs from the program across all test files. This output can then be sent to the SFS program conmat which is a general purpose program for producing confusion matrices. The script is as follows:

```
# doperf.sh
(for f in `cat test.lst`
do
   g=`echo $f|sed s/.dat//`
   ancomp -r an.02 -t an.03 -f -m - $g.sfs
done) | conmat –esl
```

If we run this on all the test files, we get this overall performance assessment:

```
$ sh doperf.sh
```

```
Processing date     : Mon Jun 28 12:44:05 2004
Confusion data from : stdin

   Confusion Matrix

      | SIL   UNV     VOI
-----+-------------------------
SIL  | 8130    855     151   9136 total  88%
UNV|  888    5010   1631    7529 total  66%
VOI |   736    863 12298  13897 total  88%

Number of matches = 30562
Recognition rate  =  83.2%
```

A way of diagnosing where the recognition is failing is to look at the mapping from the original acoustic annotations to the recognised SIL-VOI-UNV labels. We can do this for a single file as follows:

```
$ ancomp -r an.01 -t an.03 -f mac.0001.sfs
              SIL       UNV       VOI
   #:         12        1         0
  ##:        903        2         0
   +:         10        7         0
   /:          0        0         0
  3::          0        0        28
 3:?:          0        0         4
  =n:         15        6        25
 =nf:          0        0         1
   @:          1        6       144
  @?:          2        7        15
 @U@:          0        1        28
 @UU:          0        0        18
  @f:          0        8        11
  @~:          0        1        65
  A::          0        1        54
 A:f:          0        0         1
 A:~:          0        0         3
   D:         11        5        14
...
```

We could study this output to see whether we had made mistakes in our mapping from acoustic labels to classes.

Our little project could be extended in a number of ways:

1. **Use a different acoustic feature set**. A popular spectral envelope feature set is mel-scaled cepstral coefficients (MFCCs). These can be calculated with the SFS program `mfcc`.

2. **Use a different HMM configuration**. It is possible that a 3-state rather than 1-state HMM would perform better on this task, although care would have to be taken that it was still capable of recognising segments which were shorter than 3 spectral frames.

3. **Use a different density function**. Since we are mapping a wide range of spectral vectors to a few classes, it is likely that the spectral density function for each class will not be normally distributed. The use of Gaussian mixtures within the HMMs may help.

4. **Use of full covariance**. The 19 channels of the filterbank have a significant degree of covariation, which may affect the accuracy of the probability estimates from the HMM. It may be better to use a full covariance matrix in the HMM rather than just a diagonal covariance.

5. **Use of symbol sequence constraints**. Although unlikely to make much of an impact in this application, in many tasks there are constraints on the likely sequences of recognised symbols. HTK allows us to put estimated sequence probabilities in the recognition network.

3. Phone recognition

In this section we will build a simple phone recogniser. Again the aim is not to get ultimate performance, but to demonstrate the steps and the tools involved.

Source Data

Training a phone recogniser requires a lot of data. For a speaker-dependent system you need several hundred sentences, while for a speaker-independent system you need several thousand. For this tutorial we will use the WSJCAM0 database. This is a database of British English recordings modeled after the Wall Street Journal database. The WSJCAM0 database is available from the Linguistic Data Consortium (www.ldc.upenn.edu). This database is large and comes with a phone labeling that makes it very easy to train a phone recogniser.

For the purposes of this tutorial we will just use a part of the speaker-independent training set. We will use speakers C02 to C0Z for training, and speakers C10 to C19 for testing. Within each speaker we only use the WSJ sentences, which are coded with a letter 'C' in the fourth character of the filename. The first thing to do is to obtain a list of file 'basenames' - just the directory name and basename of each file we will use for training and for testing. The following script will do the job:

```
# dogetnames.sh – get basenames of files for training and testing
rm -f basetrain.lst
for d in c:/data/wsjcam0/si_tr/C0*
do
  echo processing $d
  for f in $d/???C*.PHN
  do
    g=`echo $f | sed s/.PHN//`
    if test -e $g.WV1
    then
       h=`echo $g | sed s%c:/data/wsjcam0/si_tr/%%`
        echo $h >>basetrain.lst
    fi
  done
done
rm -f basetest.lst
for d in c:/data/wsjcam0/si_tr/C1[0-9]
do
  echo processing $d
  for f in $d/???C*.PHN
  do
    g=`echo $f | sed s/.PHN//`
    if test -e $g.WV1
    then
       h=`echo $g | sed s%c:/data/wsjcam0/si_tr/%%`
       echo $h >>basetest.lst
    fi
  done
done
```

The script looks for phonetic annotation files of the form "???C*.PHN" and as long as there is a matching audio signal ".WV1" it adds the basename of the file to a list. This script creates a file called "basetrain.lst"; which has the speaker directory and base filename for each training file, and a file called "basetest.lst", which has the speaker directory and base filename for each test file. For the data used, there are about 3000 files in basetrain.lst and 900 in basetest.lst.

We can now load the audio signal and the source phonetic annotations into an SFS file using the following script:

```
# domakesfs.sh
#
# 1. Make training and testing directories
#
mkdir train test
for d in 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
do
```

```
  mkdir train/C0$d
done
for d in 0 1 2 3 4 5 6 7 8 9
do
  mkdir test/C1$d
done
#
# 2. Convert audio and labels to SFS
#
for f in `cat basetrain.lst`
do

  cnv2sfs c:/data/wsjcam0/si_tr/$f.wv1 train/$f.sfs
  anload -f 16000 -s c:/data/wsjcam0/si_tr/$f.phn train/$f.sfs
done
for f in `cat basetest.lst`
do
  hed -n test/$f.sfs
  cnv2sfs c:/data/wsjcam0/si_tr/$f.wv1 test/$f.sfs
  anload -f 16000 -s c:/data/wsjcam0/si_tr/$f.phn test/$f.sfs
done
```

This script decompresses the audio signals in the ".WV1" files and loads in the phonetic annotations. The SFS files are created in "train" and "test" subdirectories of the tutorial folder:

$ mkdir tutorial2
$ cd tutorial2
$ sh dogetnames.sh
$ sh domakesfs.sh

Figure E.3.1 shows an example source data file with an audio signal and phonetic annotations.

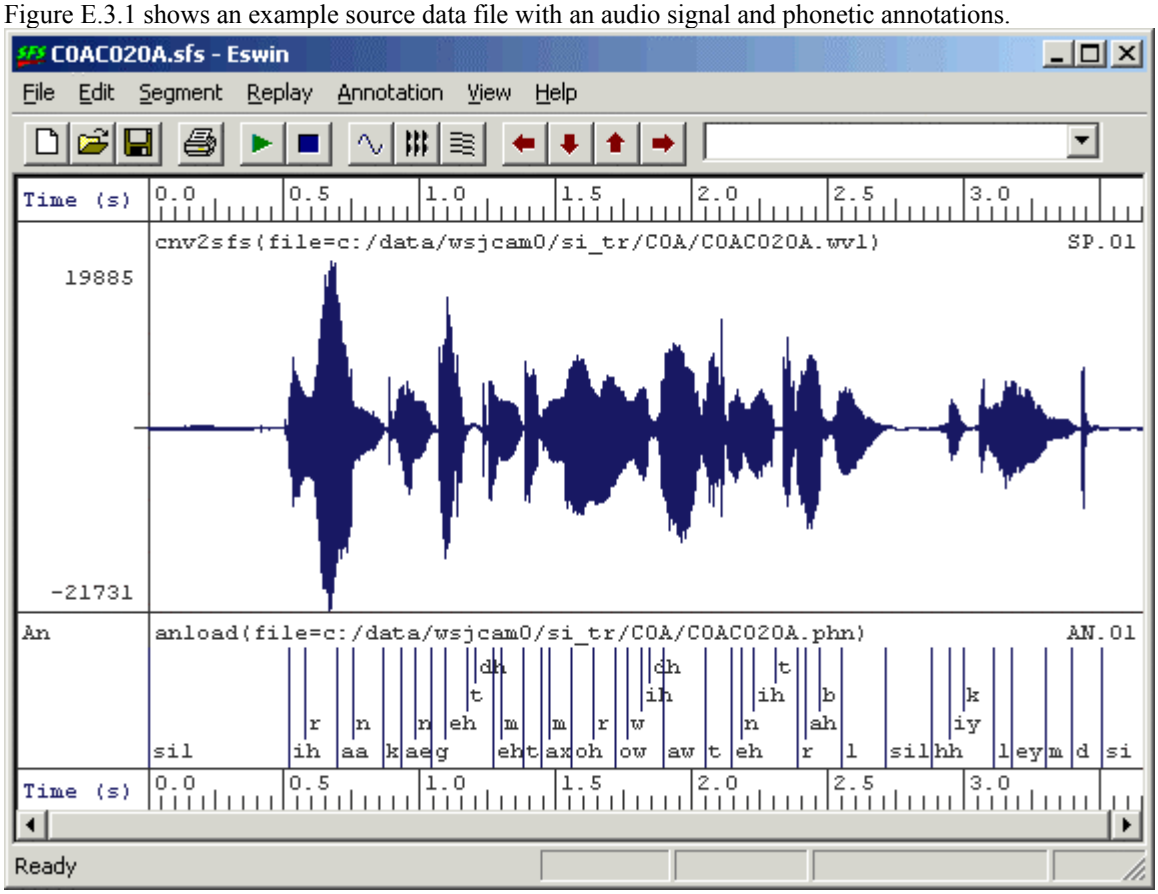


Figure E.3.1 - Source data from WSJCAM0

Making HTK data files

We can now choose an acoustic feature set for the data and generate suitable HTK format data files and label files.

A very common type of acoustic feature set for phone recognition is based on mel-scaled cepstral coefficients (MFCCs). To keep things simple, we will use 12 MFCC parameters plus one parameter that is the log energy for each 10ms frame of signal. We will use the SFS program mfcc for this, but in fact HTK has its own tool for calculating MFCCs. Also for speed we will not use delta or delta-delta coefficients although these have been shown to improve recognition performance in some circumstances. The following shell script performs the MFCC calculation, saves the data into an HTK format file and records the HTK file names in 'train.lst' and 'test.lst' for us to use when we are training and testing HMMs.

```
# domakedat.sh
#
rm -f train.lst
for f in `cat basetrain.lst`
do
  mfcc -n12 -e -l100 -h6000 train/$f.sfs
  colist -H train/$f.sfs
  echo train/$f.dat >>train.lst
done
rm -f test.lst
for f in `cat basetest.lst`
do
  mfcc -n12 -e -l100 -h6000 test/$f.sfs
  colist -H test/$f.sfs
  echo test/$f.dat >>test.lst
done
```

To save producing one HTK label file per data file, we will create an HTK "Master Label File", which will hold all the phone labels for all files. Master label files can be built using the HTK HLEd program, but here we will use an SML script. This is easier to run and allows us to collect a list of the phone names and build a phone dictionary at the same time. The SML script is as follows:

```
/* makemlf.sml – make HTK MLF file from files */

/* table to hold annotation labels */
string    table[1:1000];
var    tcount;

/* MLF file */
file    op;

/* function to check/add label */
function var checklabel(str)
{
   string str;

   if (entry(str,table)) return(0);
   tcount=tcount+1;
   table[tcount]=str;
   return(1);
}

/*initialise */
init {
   openout(op,"phone.mlf");
   print#op "#!MLF!#\n";
}

/* for each input file */
main {
   var        i,num;
```

```
    string     basename;
    string    label;

    /* print filename */
    print $filename,"\n"
    i=index("\.",$filename);
    if (i) basename=$filename:1:i-1 else basename=$filename;
    print#op "\"",basename,".lab\"\n";

    /* print annotations */
    num=numberof(".");
    for (i=1;i<=num;i=i+1) {
       label = matchn(".",i);
       print#op label,"\n";
       checklabel(label);
    }
    print#op ".\n"
}

/* output phone list and dictionary */
summary {
    var    i,j;
    string    t;

    /* insertion sort */
    for (i=2;i<=tcount;i=i+1) {
       j=i;
       t=table[j];
       while (compare(t,table[j-1])<0) {
          table[j] = table[j-1];
          j=j-1;
          if (j==1) break;
       }
       table[j]=t;
    }

    /* close MLF file */
    close(op);

    /* write phone list */
    openout(op,"phone.lst");
    for (i=1;i<=tcount;i=i+1) print#op table[i],"\n";
    close(op);

    /* write phone+ list */
    openout(op,"phone+.lst");
    print#op "!ENTER\n";
    print#op "!EXIT\n";
    for (i=1;i<=tcount;i=i+1) print#op table[i],"\n";
    close(op);

    /* write phone dictionary */
    openout(op,"phone.dic");
    print#op "!ENTER []\n";
    print#op "!EXIT []\n";
    for (i=1;i<=tcount;i=i+1) print#op table[i],"\t",table[i],"\n";
    close(op);
}
```

This script is run as follows:

$ sml -f makemlf.sml train test

The "-f" switch to SML means that it ignores non SFS files when it searches directories and sub-directories for files. The output is "phone.mlf" - the HTK master label file, "phone.lst", a list of the phone labels used in the data, "phone+.lst", a list of the phones augmented with enter and exit labels, and "phone.dic", a dictionary in which phone "word" symbols are mapped to phone pronunciations (see section 2 for explanation!).

Training HMMs

To start we will need an HTK global configuration file just as we had in section E.2. Here it is again - put this in "config.txt":

```
# config.txt - HTK basic parameters
SOURCEFORMAT = HTK
TARGETKIND = MFCC_E
NATURALREADORDER = T
```

We are now in a position to train a set of phone HMMs, one model per phone type. We will construct these in a fairly conventional way with 3 states and no skips, with one gaussian mixture of 13 dimensions per state. Put the following in a file "proto-3-13.hmm":

```
<BeginHMM>
<NumStates> 5 <VecSize> 13 <MFCC_E>
 <State> 2
  <Mean> 13
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 13
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 3
  <Mean> 13
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 13
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 4
  <Mean> 13
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 13
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <TransP> 5
  0.0 1.0 0.0 0.0 0.0
  0.0 0.6 0.4 0.0 0.0
  0.0 0.0 0.6 0.4 0.0
  0.0 0.0 0.0 0.7 0.3
  0.0 0.0 0.0 0.0 1.0
<EndHMM>
```

We now initialise this model to be "flat" - that is with each state mean set to the global data mean and each state variance set to the global data variance. The HTK tool HCompV will do this, as follows:

$ HCompV -T 1 -C config.txt -m -S train.lst -o proto.hmm proto-3-13.hmm

This creates a file "proto.hmm" with the <MEAN> and <VARIANCE> sections initialised to appropriate values. We now need to duplicate this prototype into a model for each phone symbol type. We can do this with a shell script as follows:

```
# domakehmm.sh
HCompV -T 1 -C config.txt -m -S train.lst -o proto.hmm proto-3-13.hmm
Head -3 proto.hmm > hmmdefs
for s in `cat phone.lst`
do
  echo "~h \"$s\"" >>hmmdefs
  gawk '/BEGINHMM/,/ENDHMM/ { print $0 }' proto.hmm >>hmmdefs
done
```

Run this script to creates a file "hmmdefs" which has a model entry for each phone all initialised to the same mean values.

We can now train the phone models using the HTK embedded re-estimation tool HERest. The basic command is as follows:

$ HERest -C config.txt -I phone.mlf -S train.lst -H hmmdefs phone.lst

This command re-estimates the HMM parameters in the file hmmdefs, returning the updated model to the same file. This command needs to be run several times, as the re-estimation process is iterative. To decide how many cycles of re-estimation to perform it is usual to monitor the performance of the recogniser as it trains and to stop training when performance peaks. For this to work we need to test the recogniser on material that hasn't been used for training, and for honesty won't be used for the final performance evaluation either.

To estimate the performance of the recogniser, we use it to recognise our reserved test data and compare the recognised transcriptions to the ones distributed with the database. To perform recognition we need the phone.lst file, which lists the names of the models, the phone.dic file, which maps the phone names onto themselves, and a phone.net file, which contains the recognition grammar. In this application it makes sense to use a bigram grammar in which we record the probabilities that one phone can follow another. We can build this with the commands:

```
$ HLStats -T 1 -C config.txt -b phone.big -o phone.lst phone.mlf
$ HBuild -T 1 -C config.txt -n phone.big phone+.lst phone.net
```

The HLStats command collects bigram statistics from our master label file and stores them in phone.big. The HBuild command converts these into a network grammar suitable for recognition. The phone+.lst file is the list of phone models augmented with the symbols "!ENTER" and "!EXIT".

The basic recognition command, now, is just:

```
$ HVite -T 1 -C config.txt -H hmmdefs -S test.lst -i recout.mlf \
-w phone.net phone.dic phone.lst
```

This command runs the recogniser and stores its recognition output in the recout.mlf master label file. To compare the recognised labels to the distributed labels, we can use the HResults program:

```
$ HResults -I phone.mlf phone.lst recout.mlf
====================== HTK Results Analysis =======================
  Date: Thu Jul  1 09:10:58 2004
  Ref : phone.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=903, N=903]
WORD: %Corr=48.03, Acc=41.52 [H=31884, D=10898, S=23605, I=4319, N=66387]
===================================================================
```

We can put this all together in a script which runs through 10 cycles of re-estimation and collects the performance after each cycle in a log file:

```
# dotrainrec.sh
rm -f log
for n in 1 2 3 4 5 6 7 8 9 10
do
  HERest -T 1 -C config.txt -I phone.mlf -S train.lst -H hmmdefs phone.lst
  Hvite -T 1 -C config.txt -H hmmdefs -S test.lst -i recout.mlf \
  -w phone.net phone.dic phone.lst
  echo "Cycle $n:" >>log
  Hresults -I phone.mlf phone.lst recout.mlf >>log
Done
```

Figure E.3.2 shows how the Accuracy figure changes with training cycle on our data:

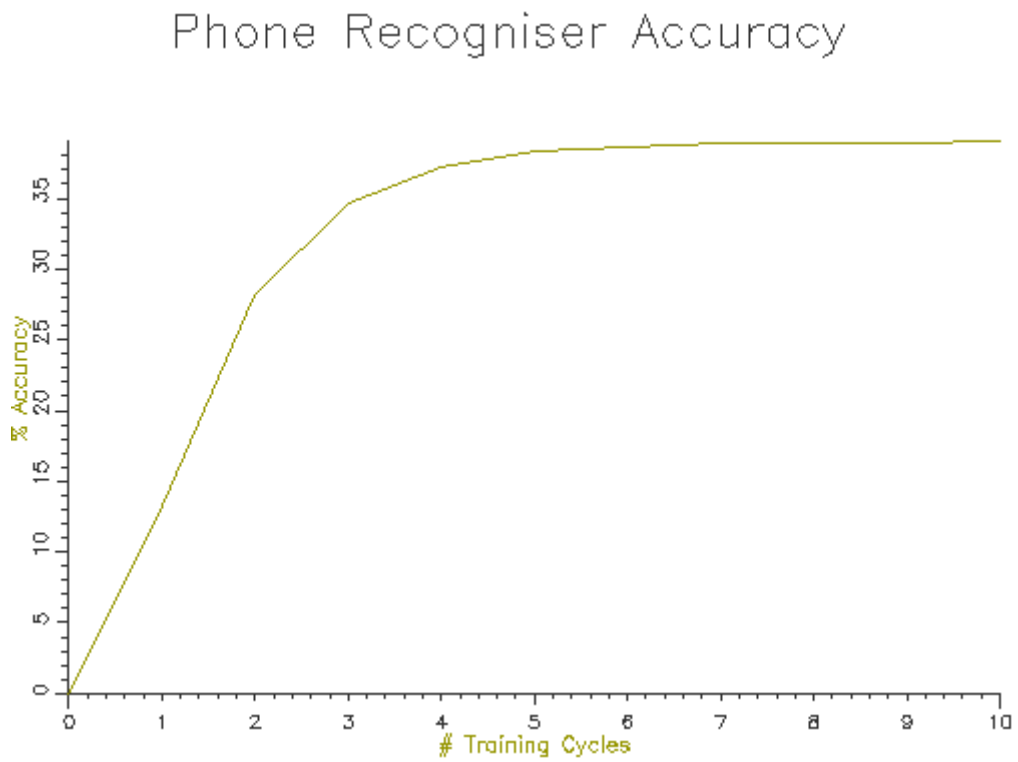


Figure E.3.2 - Phone recognition accuracy with number of training cycles

Our recogniser seems to reach a maximum performance of about 40% phone accuracy. This is not very good; the best phone recognisers on this data have an accuracy of 70%.

Parameter Analysis

One possible contributory factor to the relatively poor performance of our phone recogniser might be the use of single gaussian (normal) distributions for the modelling of the cepstral coefficient variation within a state. We can easily write an SML script to investigate whether the actual distributions are not modeled well by a gaussian distribution. The following script asks for a segment label and then scans the source SFS files to build histograms of the first 12 cepstral coefficients as they vary within instances of that segment. The distributions are plotted with an overlay of a normal distribution.

```
/* codist.sml - plot distributions of MFCC data values */

/* raw data */
var    rdata[12,100000];
var    rcount;

/* distributions */
stat    rst[12];

/* segment label to analyse */
string    label;

/* graphics output */
file gop;

/* normal distribution */
function var normal(st,x)
stat st;
{
   var x;
   x = x - st.mean;
   return(exp(-0.5*x*x/st.variance)/sqrt(2*3.14159*st.variance));
```

```
}

/* plot histogram overlaid with normal distribution */
function var plotdist(gno,st,tab,tcnt)
stat st;
var tab[];
{
   var gno;
   var tcnt;
   var i,j,nbins,bsize;
   var hist[0:100];
   var xdata[1:2];
   var ydata[0:10000];

   /* find maximum and minimum in table */
   xdata[1]=tab[gno,1];
   xdata[2]=tab[gno,1];
   for (i=2;i<=tcnt;i=i+1) {
      if (tab[gno,i] < xdata[1]) xdata[1]=tab[gno,i];
      if (tab[gno,i] > xdata[2]) xdata[2]=tab[gno,i];
   }

   /* set up x-axes */
   plotxdata(xdata,1)

   /* estimate bin size */
   nbins = sqrt(tcnt);
   if (nbins > 100) nbins=100;
   bsize = (xdata[2]-xdata[1])/nbins;

   /* calculate histogram */
   for (i=1;i<=tcnt;i=i+1) {
      j=trunc((tab[gno,i]-xdata[1])/bsize);
      hist[j]=hist[j]+1/tcnt;
   }

   /* plot histogram */
   plotparam("title=C"++istr(gno));
   plotparam("type=hist");
   plot(gop,gno,hist,nbins);

   /* plot normal distribution */
   plotparam("type=line");
   for (i=0;i<=10*nbins;i=i+1) ydata[i]=bsize*normal(st,(xdata[1]+i*bsize/10));
   plot(gop,gno,ydata,10*nbins);
}

/* get segment name */
init {
   print#stderr "For segment : ";
   input label;
}

/* for each input file */
main {
   var i,j,num
   var    t,et;

   if (rcount >= 100000) break;

   num=numberof(label);
```

```
    for (i=1;i<=num;i=i+1) {
        t = next(CO,timen(label,i));
        et = t + lengthn(label,i);
        while (t < et) {
            if (rcount >= 100000) break;
            rcount=rcount+1;
            for (j=1;j<=12;j=j+1) {
                rdata[j,rcount] = co(4+j,t);
                rst[j] += co(4+j,t);
            }
            t = next(CO,t);
        }
    }
}

/* plot */
summary {
    var j;

    openout(gop,"|dig -g -s 500x375 -o dig.gif");
    plottitle(gop,"MFCC Distributions for /"++label++"/");
    plotparam("horizontal=4");
    plotparam("vertical=3");

    for (j=1;j<=12;j=j+1) plotdist(j,rst[j],rdata,rcount);
}
```

This script can be run as

$ sml -f codist.sml train

For segment : r

Two outputs of the script are shown in Figures E.3.3 and E.3.4.

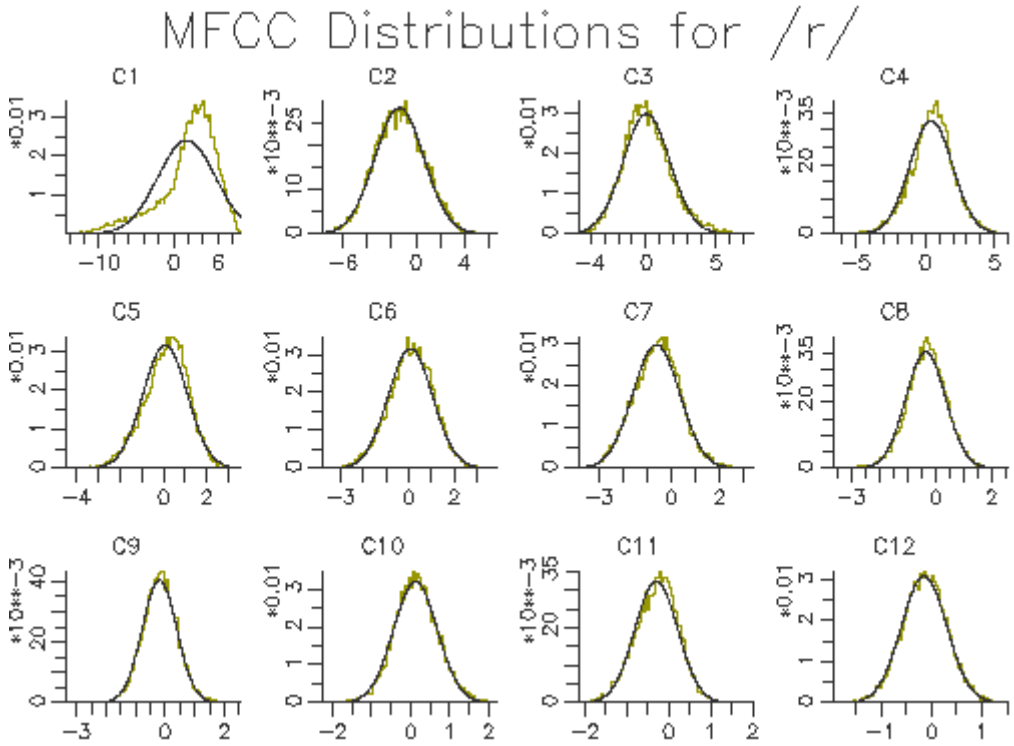


Figure E.3.3 - Modelled cepstral distributions for /r/

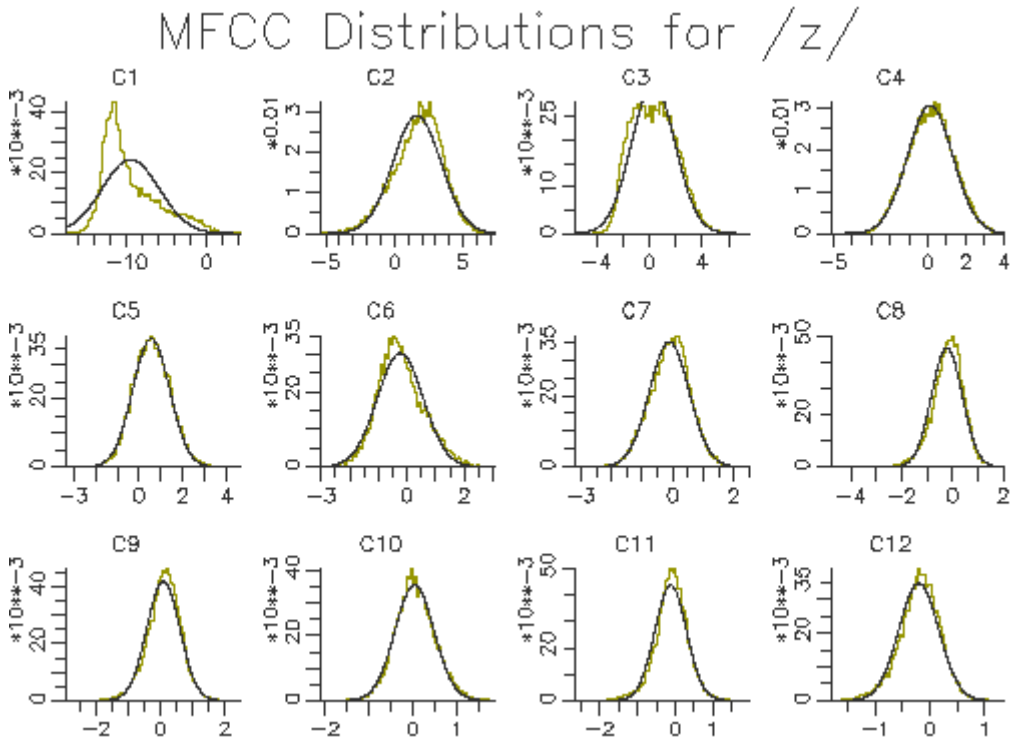


Figure E.3.4 - Modelled cepstral distributions for /z/

The figures show that a single gaussian distribution is quite good for most of the cepstral coefficients, but not for the first cepstral coefficient. This implies that we may get a small performance improvement by changing to more than one gaussian per state. To increase the number of gaussian mixtures on all phone models (for all cepstral coefficients) we can use the HTK program HHed. This program is a general purpose HMM editor and takes as input a control file of commands. In this case we just want to increase the number of mixtures on all states. Put this in a file called mix2.hed:

```
MU 2 {*.state[2-4].mix}
```

Then the HMMs can be edited with the command

```
$ HHed -H hmmdefs mix2.hed phone.lst
```

A few more cycles of training can now be applied to see the effect.

Viewing recognition results in SFS

The output of the phone recogniser above is an HTK master label file recout.mlf. If we want to view these results within SFS we need to load these as annotations. We can do this directly with the SFS program anload. First we take a copy of a test file, then load in the annotations corresponding to that file from the master label file:

```
$ cp test/c10/c10c020v.sfs .
$ anload -H recout.mlf test/c10/c10c020v.rec c10c020v.sfs
$ eswin -isp -aan c10c020v.sfs
```

Notice that the anload program takes the name of the MLF file and the name of the section in the file to load. The SFS program eswin displays the speech and annotations in the file, as shown in Figure E.3.5.

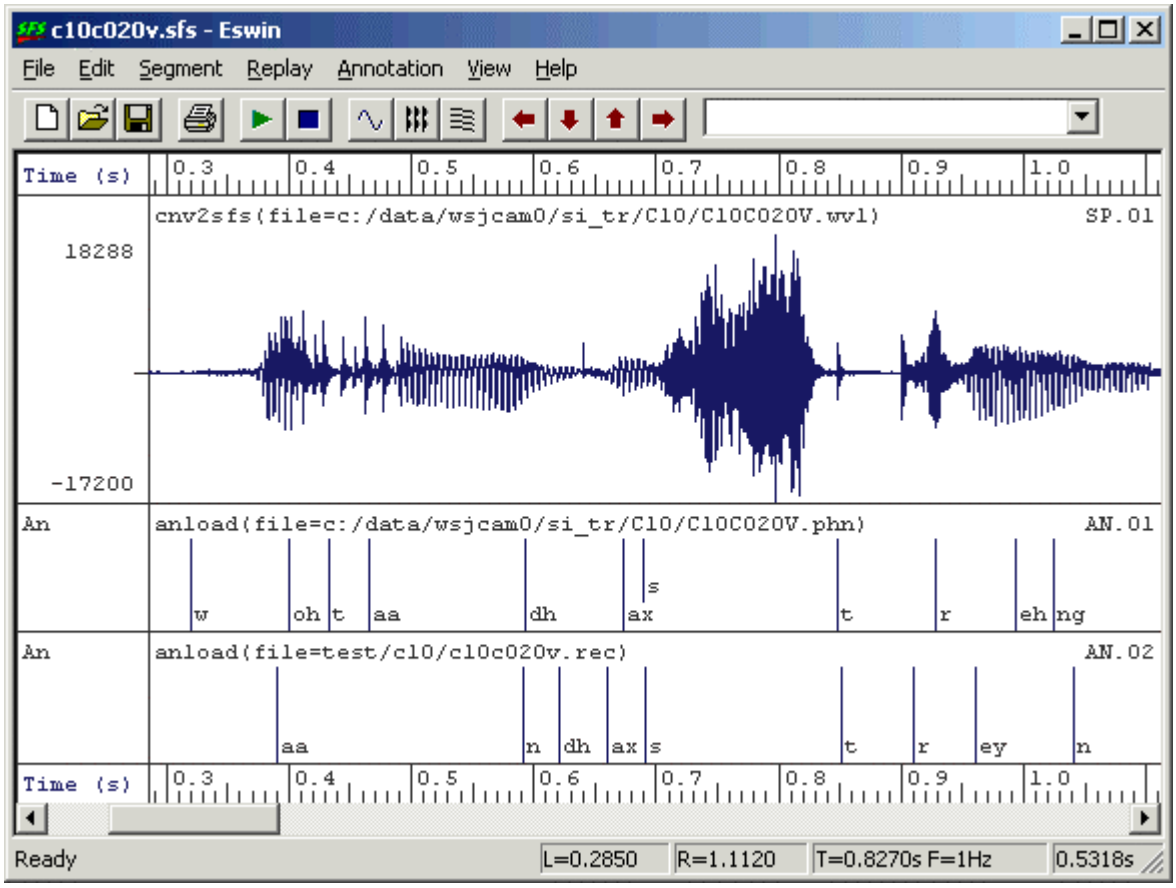


Figure.E.3.5 - Viewing recognition results

Although the HTK program HResults can analyse the performance of our recogniser on the test data and the confusions it makes, we can also perform a similar analysis in SFS for single or multiple files. The process is to load the recognised phone labels into SFS and then use the ancomp program in its "labeling" mode. We can do this on a single file with:

```
$ cp test/c10/c10c020v.sfs .
$ anload -H recout.mlf test/c10/c10c020v.rec c10c020v.sfs
$ ancomp -l c10c020v.sfs
Subst=21 Delete=6 Insert=5 Total=64 Accuracy=59.4%
```

To compare multiple files, we get ancomp to list its raw phone alignments to a file and then input the collected output to the confusion matrix program. Here is a script to do this:

```
# doancomp.sh
#
# collect mappings
(for f in `cat basetest.lst`
do
   cp test/$f.sfs temp.sfs
   anload -H recout.mlf test/$f.rec temp.sfs
   ancomp -l -m - temp.sfs
done) >ancomp.lst
rm temp.sfs
#
# build confusion matrix
conmat -esl ancomp.lst >conmat.lst
```

The file conmat.lst contains a large phoneme confusion matrix as well as an overall performance score. Since this is rather unwieldy, it is also interesting just to find the most common confusions. We can generate a list of confusions from ancomp.lst using some unix trickery:

```
$ gawk '{ if ( $1 != $2 ) print $0 }' ancomp.lst | sort | uniq -c | \
  sort -rn | head -20
    882 [] sil
```

820 ax []
724 t []
616 ih []
468 ih ax
415 n m
409 ih iy
373 ih uw
373 d []
367 n []
325 s z
313 l []
289 r []
283 t s
278 n ng
268 dh []
266 [] d
264 l ao
262 ax ih
262 ax ah

The most common confusions are probably what we would expect: insertions of silence, deletions of /@/, /t/ and /I/, substitutions of /I/ with /@/, or /m/ for /n/, and so on. How this command works is left as an exercise for the reader!

Enhancements

We might improve the performance of the phone recogniser in a number of ways:

- **Add delta coefficients**: Adding the rate of change of each acoustic parameter to the feature set (deltas) has been shown to improve phone recognition performance as has the addition of accelerations (delta-deltas). You can do this easily by changing the type of the HMM to `MFCC_E_D_A`.

- **Build phone in context models**: Building phone models which are different according to the context in which they occur can help a great deal. A typical approach to this is to build "triphone" models - where we build a model for each phone for every pair of possible left and right phones in the label files. Since this requires more data than we have typically, it is also necessary to smooth the probability estimates arising from training by "state-tying" - using the data over many triphone contexts to estimate observations for a state of one triphone model. This procedure is usually performed in a data-driven way using clustering methods. The HTK documentation gives details.

- **Add a phone language model**: Although our recogniser uses bigram probabilities to constrain recognition, there are also useful constraints on sequences longer than two phones that would help recognition. HTK does not have a simple way of doing this, but one might expect that 3-gram, 4-gram or even 5-gram phone grammars would help considerably.

4. Word recognition

Once we have built a phone recogniser, it is a simple matter to extend it to recognise words. Of course it is also possible to build a word recogniser in which each word is modelled separately with an HMM.

Dictionary

To build a word recogniser from a phone recogniser, we first need a dictionary that maps words to phone sequences. Here is an example for a simple application. Put this in digits.dic:

```
ZERO        z ia r ow
ZERO        ow
ONE         w ah n
TWO         t uw
THREE       th r iy
FOUR        f ao
FOUR        f ao r
```

```
FIVE          f ay v
SIX           s ih k s
SEVEN         s eh v n
EIGHT         ey t
NINE          n ay n
WHAT-IS       w oh t ih z
PLUS          p l ah s
MINUS         m ay n ax s
TIMES         t ay m z
DIVIDED-BY    d ih v ay d ih d b ay
SIL    []     sil
```

Grammar

Next we need a grammar file which constrains the allowed word order. The more constraints we can put here, the more accurate our recogniser. Here is a simple grammar file that allows us to recognise phrases such as "what is two plus five". Put this in digits.grm:

```
$digit = ONE | TWO | THREE | FOUR | FIVE | SIX | SEVEN | EIGHT | NINE | ZERO;
$operation = PLUS | MINUS | TIMES | DIVIDED-BY;
( SIL WHAT-IS <$digit> $operation <$digit> SIL )
```

To convert this file to a form that the recogniser can use, we need to run the HTK tool HParse, as follows:

```
$ HParse digits.grm digits.net
```

Word recogniser

The basic command for recognising HTK data file inp.dat using our digits task recogniser is then just:

```
$ HVite -T 1 -C config.txt -H hmmdefs -w digits.net digits.dic \
  phone.lst inp.dat
```

We can put together a simple script that uses SFS to acquire an audio signal, then performs an MFCC analysis, exports the coefficients to HTK and runs the recogniser:

```
# doreclive.sh
rm -f inp.sfs
hed -n inp.sfs >/dev/null
echo "To STOP this script, type CTRL/C"
#
remove -e inp.sfs >NUL
echo "***** Say Word *****"
while record -q -e -f 16000 inp.sfs
do
  replay inp.sfs
  mfcc -n12 -e -l100 -h6000 inp.sfs
  colist -H inp.sfs
  HVite -T 1 -C config.txt -H hmmdefs -w digits.net digits.dic \
    phone.lst inp.dat
  remove -e inp.sfs >NUL
  echo "***** Say Word *****"
done
```

You could also use the HTK live audio input facility for this demonstration. But then you should also use the HTK MFCC analysis in training as well, as the HTK MFCC analysis gives slightly different scaled coefficient values from the SFS program mfcc.

Enhancement

We might enhance our word recogniser in a number of ways:

- Use a better set of phone models, for example with triphones.
- For a small vocabulary and lots of training data, it is better to build word-level models.

- For a large vocabulary it is better to use a statistical language model, such as a trigram model rather than trying to build a grammar.

- Model non-speech regions and silent gaps more intelligently, by having models for noises and inter-word silent gaps for example.

5. Phone alignment

Another application for our phone recogniser is phone alignment. In phone alignment we have an unlabeled audio signal and a transcription and the task is to align the transcription to the signal. This procedure is already implemented as part of SFS using the program analign, but we will show the basic operation of analign here.

Assume we have an audio recording of a single sentence, here it is the sentence "six plus three equals nine" stored in an sfs file called six.sfs. We first perform MFCC analysis on the audio signal:

```
$ mfcc -n12 -e -l100 -h6000 six.sfs
```

We next add an annotation containing the raw transcription in ARPABET format:

```
$ anload -t phone -T "sil s ih k s p l ah s th r iy iy k w ax l z n ay n sil" six.sfs
```

Then export both coefficients and annotations to HTK format:

```
$ colist -H six.sfs
$ anlist -h -O six.sfs
```

We can now run the HTK HVite program in alignment mode with:

```
$ HVite -C config.txt -a -o SM -H hmmdefs phone.dic phone.lst six.dat
```

And load the aligned annotations back into SFS:

```
$ anload -h six.rec six.sfs
```

The result is shown in figure E.5.1.

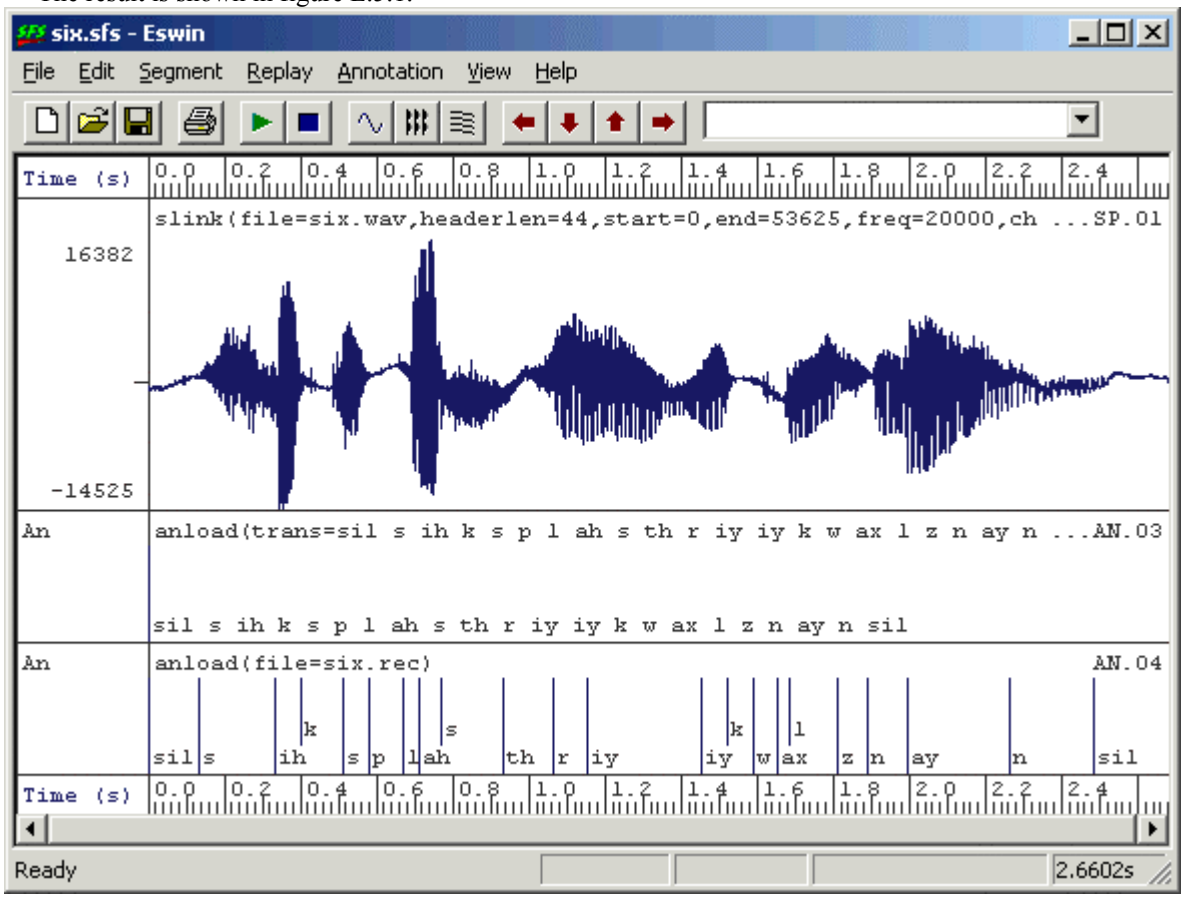


Figure E.5.1 - Aligned phone labels

The SFS program analign makes this process easier by handling all the export and import of data to HTK and also handles the chopping of large files into sentence sized pieces and the translation of symbol sets between SAMPA, ARPABET and JSRU.

6. Pronunciation variation analysis

A variation on phone alignment is to provide a shallow network of pronunciation alternatives to HVite and let it choose which pronunciation was actually used by a speaker. This kind of variation analysis could be useful in the study of accent variation or when attempting to recognized dysfluenceies from speakers with different accents.

In this example, we will take a sentence spoken by two speakers with different regional accents of the British Isles. The sentence is "after tea father fed the cat", and we are interested whether the vowel /{/ or /A:/ was used in the words "after" and "father". For reference, we always expect "cat" to be produced with /{/, so we will check on the analysis by also looking to see if the program rejects the pronunciation of "cat" as /kA:t/.

We first create a dictionary file for the sentence. Put this in accents.dic:

```
AFTER      aa f t ax
AFTER      ae f t ax
TEA        t iy
FATHER     f aa dh ax
FATHER     f ae dh ax
FED        f eh d
THE        dh ax
CAT        k aa t
CAT        k ae t
SIL    []  sil
```

Notice that the dictionary lists the alternative pronunciations in which we are interested. Next we create a grammar just for this sentence. Put this in accents.grm

```
( SIL AFTER TEA FATHER FED THE CAT SIL )
```

Next we convert the grammar file to a recognition network with HParse:

```
$ HParse accents.grm accents.net
```

Then we prepare HTK data files for the two audio signals:

```
$ mfcc -n12 -e -l100 -h6000 brm.sfs
$ mfcc -n12 -e -l100 -h6000 sse.sfs
$ colist -H brm.sfs
$ colist -H sse.sfs
```

Then we recognise the utterances using the grammar, but outputting the selected phone transcription.

```
$ HVite -C config.txt -H hmmdefs -w accents.net -m -o ST accents.dic \
  phone.lst brm.dat sse.dat
```

The "-m" switch to HVite causes the phone model names to be output to the recognised label files, while the "-o ST" switch cause the scores and the times to be suppressed. The output of the recogniser are in brm.rec and sse.rec as follows:

brm.rec	**sse.rec**
sil SIL	sil SIL
ae AFTER	aa AFTER
f	f
t	t
ax	ax
t TEA	t TEA
iy	iy
f FATHER	f FATHER
aa	aa
dh	dh
ax	ax
f FED	f FED
eh	eh
d	d
dh THE	dh THE
ax	ax
k CAT	k CAT
ae	ae
t	t
sil SIL	sil SIL

From this it is easy to see that the speaker from Birmingham used /{/ where the speaker from South-East England used /A:/ in "after".

7. Dysfluency recognition

Another application of phone recognition is the detection of dysfluencies. Although our phone recogniser is not particularly accurate we can still look for patterns in the recognised phone sequence which may be indicators of dysfluency. We will recognise a passage with the phone recogniser, load the phone labels into the SFS file then use an SML script to look for dysfluent patterns.

Here are the commands for performing MFCC analysis on the passage, saving the coefficients to HTK format, building a simple phone loop recogniser without bigram constraints, then running the recogniser and loading the phone annotations back into the file:

```
$ mfcc -n12 -e -l 100 -h 6000 dysfluent.sfs
$ colist -H dysfluent.sfs
$ HBuild phone.lst phone.net
$ HVite -C config.txt -H hmmdefs -w phone.net -o S phone.dic phone.lst dysfluent.dat
$ anload -h dysfluent.rec dysfluent.sfs
```

The following SML script takes the recognised phone sequence and looks for (i) single (non-silent) phones lasting for longer than 0.25s, (ii) repeated (non-silent) phone labels which together last for longer than 0.25s, (iii) patterns of five phone labels matching A-B-A-B-A where A is not silence. Output is another annotation item with the dysfluent regions marked with "(D)".

```
/* dysfind.sml – find dysfluencies from phone recogniser output */

/* input and output annotation sets */
item    ian;
item    oan;

/* table to hold dysfluent events */
var    times[1000,2];
var    tcount;

/* add times of dysfluencies to table */
function var addtime(posn,size)
{
   var posn,size;
   var i;

   /* put event in sorted position */
   i=tcount+1;
   if (i > 1) {
      while (posn < times[i-1,1]) {
         times[i,1] = times[i-1,1];
         times[i,2] = times[i-1,2];
         i = i - 1;
         if (i==1) break;
      }
   }
   times[i,1]=posn;
   times[i,2]=size;
   tcount=tcount+1;
}

/* process each input file */
main {
   var    i,numf,fdur,dmin;
   var ocnt,size;
   string lab1,lab2,lab3;

   /* get input & output */
   sfsgetitem(ian,$filename,str(selectitem(AN),4,2));
   numf=sfsgetparam(ian,"numframes");
   fdur=sfsgetparam(ian,"frameduration");
```

```
sfsnewitem(oan,AN,fdur,sfsgetparam(ian,"offset"),1,numf);

/* put minimum dysfluency length = 0.25s */
dmin = 0.25/fdur;

/* look for long non-silent annotations */
tcount=0;
for (i=1;i<=numf;i=i+1) {
   size = sfsgetfield(ian,i,1);
   if (size > dmin) {
      lab1 = sfsgetstring(ian,i);
      if (compare(lab1,"sil")!=0) {
         addtime(sfsgetfield(ian,i,0),size);
      }
   }
}

/* look for patterns like AA */
for (i=2;i<=numf;i=i+1) {
   lab1 = sfsgetstring(ian,i-1);
   lab2 = sfsgetstring(ian,i);
   if (compare(lab1,lab2)==0) {
      size = sfsgetfield(ian,i-1,1)+sfsgetfield(ian,i,1);
      if (size > dmin) {
         if (compare(lab1,"sil")!=0) {
            addtime(sfsgetfield(ian,i-1,0),size);
         }
      }
   }
}

/* look for patterns like AAA */
for (i=3;i<=numf;i=i+1) {
   lab1 = sfsgetstring(ian,i-2);
   lab2 = sfsgetstring(ian,i-1);
   if (compare(lab1,lab2)==0) {
      lab3 = sfsgetstring(ian,i);
      if (compare(lab1,lab3)==0) {
         size = sfsgetfield(ian,i-2,1)+sfsgetfield(ian,i-1,1)+sfsgetfield(ian,i,1);
         if (size > 0.25) {
            if (compare(lab1,"sil")!=0) {
               addtime(sfsgetfield(ian,i,0),size);
            }
         }
      }
   }
}

/* look for patterns like ABABA */
for (i=5;i<=numf;i=i+1) {
   lab1 = sfsgetstring(ian,i-4);
   lab2 = sfsgetstring(ian,i-2);
   if ((compare(lab1,lab2)==0)&&(compare(lab1,"sil")!=0)) {
      lab3 = sfsgetstring(ian,i);
      if (compare(lab1,lab3)==0) {
         lab1 = sfsgetstring(ian,i-3);
         lab2 = sfsgetstring(ian,i-1);
         if (compare(lab1,lab2)==0) {
            size = sfsgetfield(ian,i,0)+sfsgetfield(ian,i,1) \
                                                -sfsgetfield(ian,i-4,0);
            addtime(sfsgetfield(ian,i-4,0),size);
```

```
            }
        }
    }
}

/* convert times to annotations (ignoring overlaps) */
ocnt=0;
for (i=1;i<=tcount;i=i+1) {
    ocnt=ocnt+1;
    sfssetfield(oan,ocnt,0,times[i,1]);
    sfssetfield(oan,ocnt,1,times[i,2]);
    sfssetstring(oan,ocnt,"(D)");
    if ((i<tcount)&&(times[i+1,1]>times[i,1]+times[i,2])) {
        ocnt=ocnt+1;
        sfssetfield(oan,ocnt,0,times[i,1]+times[i,2]);
        sfssetfield(oan,ocnt,1,times[i+1,1]-times[i,1]-times[i,2]);
        sfssetstring(oan,ocnt,"/");
    }
}

/* save output back to file */
sfsputitem(oan,$filename,ocnt);
}
```

This script would be run with:

$ sml -ian^anload dysfind.sml dysfluent.sfs

Figure E.7.1 shows one particular pattern of "ih sil ih sil ih" which is a real dysfluency detected by the script.

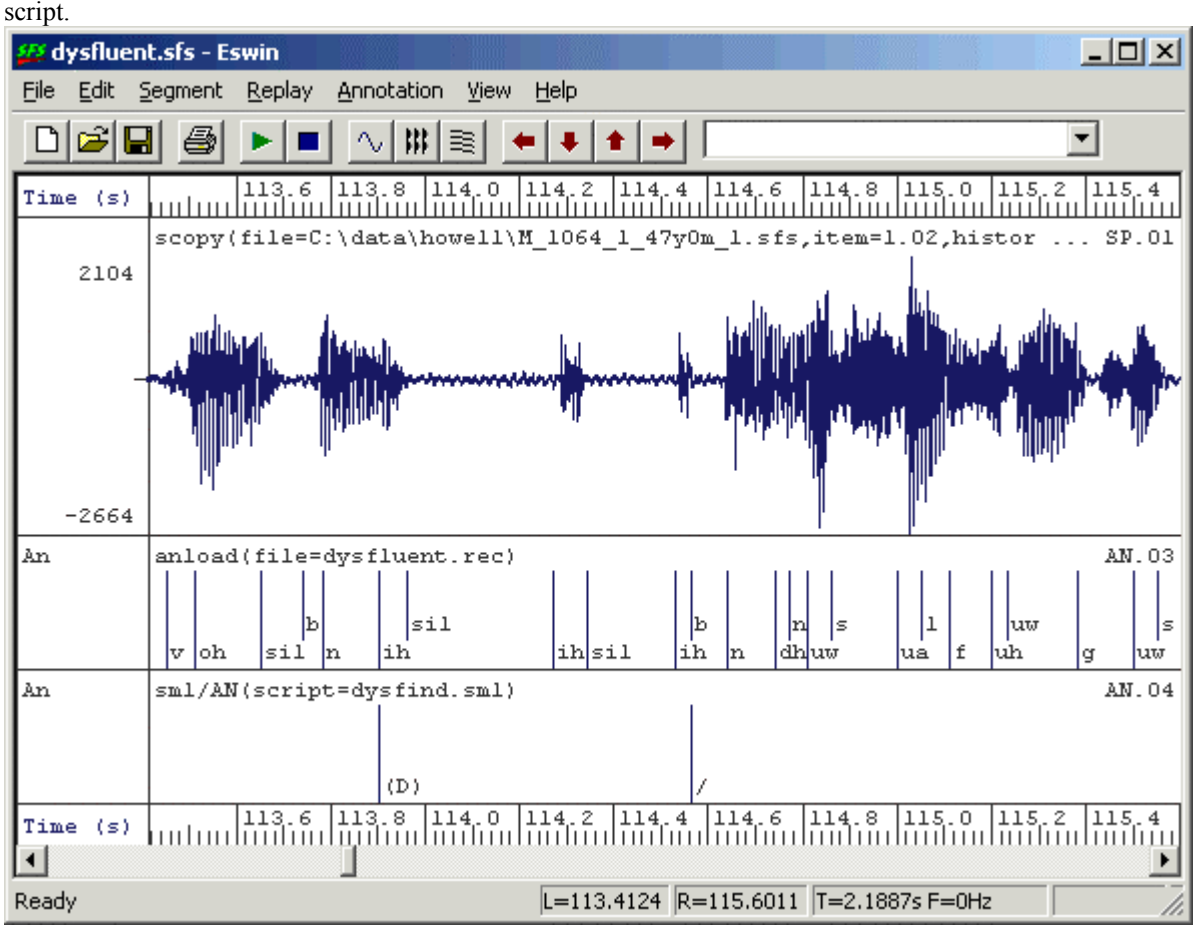


Figure E.7.1 - Dysfluency recognition

Bibliography

- BEEP British English pronunciation dictionary at ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/beep.tar.gz.
- Hidden Markov modelling toolkit at http://htk.eng.cam.ac.uk/.
- International Phonetic Alphabet at http://www.arts.gla.ac.uk/IPA/ipachart.html.
- SAMPA Phonetic Alphabet at http://www.phon.ucl.ac.uk/home/sampa/.
- Speech Filing System at http://www.phon.ucl.ac.uk/resource/sfs/.
- SCRIBE corpus at www.phon.ucl.ac.uk/resource/scribe.

Feedback

Please report errors in appendices C, D and E to SFS@phon.ucl.ac.uk. Questions about the use of SFS can be posted to the SFS speech-tools mailing list.