Exploring the phonetics of neutralisation with phonology in mind*

Mariam Almihmadi

Abstract

In this paper, I explore the phonetics of two phonological scenarios of neutralisation. In scenario A, neutralisation is phonologically complete. This applies to contrasts that are reportedly neutralised, with phonology completely giving up the underlying contrast on the surface. In scenario B, neutralisation is not phonologically complete. This refers to contrasts whose neutralisation plays a role in the creation of underapplication/overapplication opacity (Kiparsky 1973; McCarthy 1999). Until now, these scenarios have been studied separately in the experimental literature on neutralisation. Here I look closely at both, drawing on a dataset from a Hijazi dialect of Arabic that has a difference between lexical and epenthetic vowels. Results of a production experiment reveal a curious pattern of dis-correlation between the phonetics and phonology of neutralisation. I speculate on its relevance to the laboratory tradition in the study of the phonetics of neutralisation.
Keywords: incomplete neutralisation, vowel epenthesis, phonetics-phonology relation, opacity

1 Introduction

The phonetics of neutralisation has been a subject of investigation since the 70s of the last century. Contributors to the rapidly growing literature on the topic are generally concerned with establishing the extent of phonetic merger (i.e., the completeness question), and the empirical validity of the phonetic effect (i.e., the genuineness question).

Perhaps the most important finding of decades of experimentation is the discovery of incomplete neutralisation, which generally describes a situation where small yet statistically significant acoustic[1] differences exist between sounds that are reportedly neutralised. However, the genuineness of incomplete neutralisation was soon questioned mainly on the grounds that the acoustic residue was found to fluctuate to the point of disappearance under different experimental conditions (see e.g., Dmitrieva et al 2010; Jassem & Richter 1989; Port & Crawford 1989; Snoeren et al 2006; Warner et al 2004, 2006). Blame for this mostly fell on the specifics of the experimental design which, according to some, encouraged 'spelling pronunciation' (e.g., Fourakis & Iverson 1984; Warner et al 2006) and/or 'hypercorrection' (e.g., Jassem & Richter 1989). Both of those were thought to be issuing from what Barnes (2006: 234) called the 'paralinguistic contamination problem'.

Another, far less celebrated, source of scepticism concerning the genuineness of incomplete neutralisation is that the reported phonetic differences are not always in the same direction as when the relevant contrast is fully realised (e.g., Barnes 2006; Dinnsen & Charles-Luce 1984). These inconsistencies added to the controversy that was already engulfing incomplete neutralisation, especially owing to the fact that a number of early

*I thank John Harris, Moira Yip, Yi Xu, Andrew Nevins, and Neil Smith for useful comments and suggestions.

[1] Articulatory differences have also been reported in the literature (see e.g., Butcher 1995; Nolan 1992). However, for the sake of brevity, I only discuss acoustic differences here.

studies yielded mixed results even when they were investigating the same neutralisation case (see e.g., Fourakis & Iverson 1984 vs O'Dell & Port 1983 for German; Jassem & Richter 1989 vs Slowiaczek & Dinnsen 1985 for Polish; Baumann 1995 vs Warner et al 2004 for Dutch). After almost 40 years of experimentation, the basic exploratory question of the reality and mechanism of incomplete neutralisation is far from settled. Commentators on the topic continue to call for further research (e.g., Barnes 2006; Ernestus, in press).

In this paper, I focus on a much neglected side of the question—the correlation between the phonetics and phonology of neutralisation. To that end, I investigate the phonetics of two neutralisation scenarios bearing directly on the completeness question, with a view to bringing more balance to our approach to the relation between the phonetics and phonology of neutralisation. As a schematic description, scenario A, which has attracted most experimentation, involves contrasts that are impressionistically described as neutralised, with phonology completely giving up the underlying contrast on the surface. Conversely, scenario B, which is under-researched from an experimental perspective, involves contrasts which are impressionistically described as neutralised, but which phonology still refers to on the surface under certain conditions. One important condition is their role in the creation of what is known in the phonological literature as underapplication and/or overapplication opacity (e.g., Kiparsky 1973; McCarthy 1999). Until now, these scenarios have been investigated separately. But here I look more closely at both, using data from a single phonological process drawn from a single language.

Neutralisation involving vowel/zero contrasts in Bedouin Hijazi Arabic[2] (henceforth BHA) through vowel epenthesis exemplifies these scenarios. More specifically, [a]-epenthesis represents scenario A: Lexical /a/ and epenthetic [a] behave the same with respect to phonological processes. By contrast, [i]-epenthesis illustrates scenario B: Despite sounding the same to the naked ear, lexical /i/ and epenthetic [i] behave differently with respect to some phonological processes.

A phonetic investigation of vowel/zero neutralisation in BHA should provide a rare opportunity to test hypotheses involving the phonetics-phonology relation.

The rest of the paper is organised as follows: In section 2, I highlight the phonetic and phonological dimensions of the completeness question; in section 3, I give a brief description of the phonology of vowel/zero neutralisation in BHA; in section 4, I present the production experiment and analyse and discuss the findings; and in section 5, I summarise the main arguments and point out directions for future research.

2 The Completeness Question between Phonetics and Phonology

As I pointed out above, the most significant outcome of decades of phonetic inquiry into neutralisation is that neutralisation may be phonetically incomplete, with traces of the underlying contrast surviving in the acoustic signal. To many phoneticians, this discovery has strengthened calls for a greater role for phonetics in phonology. For some time it looked as though, upon the recognition of the existence of incomplete neutralisation, "[a] theoretical dilemma [for phonology]…arises concerning the presumed distinction between neutralization rules and allophonic rules", which only phonetics can resolve (Slowiaczek & Dinnsen 1985: 326) (but see Blumstein 1991 and Nolan 1986 for a different view). This justified the investment of time and effort in small-scale studies whose main mission was to prove/disprove the existence of incomplete neutralisation. Post hoc discussions, with an inherently limited scope, of the implications of these findings for the relationship between

[2] See Al-Mozainy (1981) for a description of the phonology of this dialect.

phonetics and phonology had little impact. Furthermore, their credibility was considerably undermined by the identification of a set of methodological shortcomings with the potential threat of calling into question the genuineness of the reported findings (see above for references).

To many phonologists, on the other hand, the discovery of incomplete neutralisation is only very recently beginning to assume any theoretical relevance. Incomplete neutralisation is now being taken to provide support to a number of phonological models including the Dynamics model (e.g., Gafos & Benus 2006; Nycz 2005), Candidate Chains in OT (e.g., McCarthy 2007; Gouskova & Hall 2009), Turbidity Theory (e.g., Goldrick 2000; van Oostendorp 2008), phonologisation of sound change in progress (e.g., Barnes 2006), and the recent application of exemplar-based models to neutralisation data (e.g., Ernestus & Baayen 2003; cf. Yu 2007). But before that, and apart from a total rejection by phonologists like Manaster Ramer (1996), findings of those instrumental studies on neutralisation received little attention from the phonological community. Dinnsen (1985) attributed this undue disregard on the part of phonologists to their insistence on categoricality, which led them, goes the argument, to assign priority to perception over production. Dinnsen protested that for contemporary phonologists "if for some reason production differences were discovered without the perceptual differences, those differences would be dismissed as linguistically irrelevant" (p. 273).

However, Brockhaus (1995) observed that at the time when neutralisation was becoming empirically relevant to many phoneticians in the 80's and 90's, its theoretical relevance to phonologists was declining.

Had the purpose of laboratory inquiries into neutralisation been a quest for insights from production and perception data into covert and 'problematic' phonological patterns, the "prospect of having to do phonology or phonetics in a world in which 'incomplete neutralisation' existed" might have been less "horrific", to phonologists like Manaster Ramer (1996: 514). Unfortunately, this did not happen. It was not only the approach to the production-perception relation that was biased; the very approach to the phonetics-phonology relation was also biased. It should be remembered that the null hypothesis most of those studies were testing was that phonological neutralisation is phonetically complete. What this implies for the phonetics-phonology relation is that the phonetics and phonology of neutralisation are taken to mirror each other. More specifically, phonetics is assumed to reflect the phonological surface representation, where neutralisation occurs, but be blind to the underlying representation, where contrast holds. Most researchers chose for instrumental investigation contrasts that are reportedly neutralised, where phonology completely gives up the underlying contrast on the surface. To place it within its phonological context, this would be an illustration of clause (c) in Kiparsky's (1973: 79) definition of opacity[3] given in (1). Importantly, this is the clause that depicts neutralisation as opacity. Clearly, the opacity here concerns the learning of the correct underlying representation.

(1) A process P of the form A→ B/ C__D is opaque to the extent that there are phonetic forms (i.e. surface representations) in the language having
(a) A in the environment C__D [underapplication opacity], or
(b) B derived by P in environments other than C__D [overapplication opacity], or
(c) B not derived by P in the environment C__D [neutralisation][4].

[3] For expository convenience, I follow the numbering of clauses that appears in (McCarthy 1999: 358; see also Baković 2007: 219).

[4] More concretely, a process deriving, for example, [t] from an underlying /d/ in a certain context is opaque if there exists within the language and in the same context an underlying /t/ surfacing as [t] or a [t] derived by a

Interestingly, clause (c) was only later added to the original definition that appeared in Kiparsky (1971: 621-622). However, this clause has systematically been ignored by phonologists who discuss opacity. Those who ever mention it hasten to announce its irrelevance to their discussion of the problematic, thus more interesting, cases involving the other two clauses (see e.g., McCarthy 1999; Baković 2007; Bermúdez-Otero 1999).The prescribed job for the investigator of the phonetics of a phonological neutralisation effect that conforms to clause (c), then, has been to find out if the relevant neutralisation, which is complete in the phonology, is also complete in the phonetics.

However, it is not hard to see that the correlation between the phonetics and phonology of neutralisation as assumed by those researchers could also be tested by studying a case where a phonetically incomplete neutralisation *should be* what we expect, if the hypothesis of a close correlation is to be retained. Reviewing the literature on the phonetics of neutralisation, we find that only a few studies (e.g., Gouskova & Hall 2009 for vowel epenthesis in Lebanese Arabic; Colley 2009; Fox & Terbeek 1977; Sheldon 1973; and Zue & Laferriere 1979 for flapping in American English; Ettlinger 2007 for final devoicing in Dutch) have investigated contrasts whose neutralisation plays a role in the creation of underapplication/overapplication opacity (clauses (a)/(b) in the definition in (1) above). For those cases, if the phonetics and phonology of neutralisation mirrored each other, we would expect traces of the underlying contrast to survive in the phonetics, as they did in the phonology.

The importance of testing the two possibilities for the putative correlation of the phonology and phonetics of neutralisation, courtesy of a single phonological process drawn from a single language, should be obvious. On the one hand, such an investigation would help shape and sharpen our theoretical perspective on the phonetics of neutralisation and more generally the phonetics-phonology relation. On the other, it should be empirically beneficial in the context of the genuineness question I mentioned above. Fortunately, just such a testing-ground exists and is the subject of investigation in this paper.

3 Vowel/Zero Neutralisation in BHA

In BHA, vowel epenthesis neutralises vowel/zero contrasts differently according to the quality of the inserted vowel. Specifically, the result of epenthesising [a] in a word-final CC-cluster that violates Sonority Sequencing Principle (SSP) (Clements 1990: 285) is a total obliteration of the underlying contrast on the surface. For example, [laħam] can be the output of either /laħm/ 'meat' (with epenthetic [a]) or /laħam/ 'shut tight'. Importantly, an epenthetic [a], underlyingly a zero, and a lexical /a/ are both impressionistically described and transcribed as being the same, and crucially, they behave the same with respect to phonological processes. Or to be more precise, there is no process in the phonology of BHA that treats epenthetic [a] and lexical /a/ differently. For space reasons, I limit the discussion here to only one state of affairs—stressability. Here, stressability is the capacity of a vowel to

rule other than the /d/ →[t]. In other words, in a language of this type, *only some* surface [t]s are underlyingly /t/s. Note that this view of neutralisation as involving opacity makes little sense under a broader conception of neutralisation that does not distinguish between 'the obliteration of contrasts that exist at the lexical level' and 'the static lack of contrast at the lexical level'. An important question to ask here is whether or not such a distinction is warranted (see Gurevich 2004; Hansson 2008). I leave this for future research.

carry word-stress when it occurs in a stress site as per the algorithm in (2) which describes stress assignment in BHA[5].

(2) a. Considering only the last three syllables, stress the rightmost heavy syllable:
[kita:tí:<b>] 'schools'; [miká:ti<b>] 'offices'; [ʔalmaktibá:<t>] 'the libraries' ; [máktiba<h>] 'a library'; [mákta<b>] 'a desk/an office'; [kitáb<t>] '(I/you) wrote'; [kítba<t>] 'was written (fem.)'.

b. Otherwise, stress the third rightmost light syllable:
[ʕárafa<h>] 'Arafat'; [gála<m>] 'a pen'; [ʔalħálaga<h>] 'the street market'.

Stress assignment in BHA is regular and insensitive to the quality of lexical vowels. As to epenthetics, there is a difference. More specifically, epenthesising [a] does not disrupt stress assignment. That is, both epenthetic [a] and lexical /a/ are stressable:

(3) /laħam+ha/ [laħámha] 'shut it tight'
/laħm+ha/ [laħámha] 'her meat'

In contrast, epenthesising [i] to break up a CC-cluster which, too, violates SSP, results in the same vowel/zero contrast being neutralised, but not quite. For example, [gidir] can be the output of either /gidr/ 'pot' (with epenthetic [i]) or /gadir/ 'managed/overpowered' (with independent raising of the first vowel). Yet, despite sounding the same to the naked ear, lexical and epenthetic high vowels behave differently with respect to some phonological processes. Sticking with stress assignment for illustration, we note that [i]-epenthesis renders stress assignment opaque or non-surface-true in McCarthy's (1999) sense:

(4) /gadir+ha/ [gidírha] 'overpowered her'
/gidr+ha/ [gídirha] 'her pot'

Further evidence for the non-stressability of [i] comes from the single environment where epenthetic vowels "are systematically stressed in all [Arabic] dialects", according to Farwaneh (1995: 151). This is when an epenthetic breaks up a CCCC-cluster. However, in BHA, epenthetic [i] appears in this environment unstressed, even though it seemingly occupies what should be a stress site by (2).

(5) (Relevant epenthetics boldfaced)

Dialect	Epenthetic in CCCC	Example
Makkan	stressed	[katabt**á**llaha]
Egyptian	stressed	[katabt**í**lha]
Palestinian	stressed	[katabt**í**lha]
Syrian	stressed	[katabt**í**lla]
BHA	Not stressed	[kitábt**i**lha]

[5] Like most Arabic dialects, BHA is a quantity sensitive language with final consonant extrametricality (Al-Mozainy 1981; Al-Mozainy et al 1985; Oh 1998). In the examples above, I have marked the extrametrical final C as <C>.

As we can see, the underlying vowel/zero contrast is apparently neutralised through vowel epenthesis in BHA. However, to decide whether the contrast is completely obliterated on the surface or not, a closer look into the phonological visibility of the epenthesised sound is necessary. Phonological visibility is limited here to vowel stressability. Impressionistically, both [a] and /a/ occurring in comparable phonological environments sound the same and behave the same with respect to phonological processes. In contrast, [i] and /i/ too sound the same, but are sometimes treated differentially by the phonology, in that [i] but not /i/ creates opacity when it interacts with certain phonological processes. It is an empirical question whether or not the phonetics of vowel/zero neutralisation correlates with its phonology. I set out to investigate this next.

4 The Experiment

4.1 Design

4.1.1 Participants. Seven native speakers of BHA aged between 22 and 60 (median age= 31) participated in this experiment. They are all monolingual female relatives of mine.

4.1.2 Materials. The stimulus set in this experiment is composed of fourteen [CV_1CV_2C] minimal pairs, contrasting with respect to the underlying status of their V_2, which is either epenthetic or lexical. These pairs appear in (6). They were quasi-randomized. Twenty fillers were appended to the beginning and end of the stimulus list, while at least 30 fillers were inserted between the items.

(6)	$V2_{lexical}$	$V2_{epenthetic}$
	[laħam] 'shut tight'	[laħam] 'meat'
	[naħal] 'teased'	[naħal] 'bees'
	[nahar] 'yelled at'	[nahar] 'river'
	[gahar] 'oppressed'	[gahar] 'oppression'
	[daxal] 'entered'	[daxal] 'income'
	[faħam] 'was out of breath'	[faħam] 'char coals'
	[rahan] 'pledged'	[rahan] 'mortgage'
	[naħar] 'slaughtered'	[naħar] 'the act of slaughtering'
	[ʃahar] 'scalded'	[ʃahar] 'month'
	[gidir] 'managed/overpowered'	[gidir] 'pot'
	[kibir] 'grew'	[kibir] 'conceit'
	[gabil] 'Gabil: proper name'	[gabil] 'before'
	[ðikir] 'remembered'	[ðikir] 'prayers'
	[fikir] 'came to realise'	[fikir] 'thought'

The stimulus set is balanced both phonologically and lexically. Phonologically, only words of the form [CV_1CV_2C] stressed on V_1 and containing clusters violating SSP when V_2 is underlyingly a zero were selected. Lexically, members of each pair in the stimulus set are matched for frequency. Recent research on word production has suggested that certain lowlevel acoustic effects are attributable to frequency (e.g., Munson & Solomon 2004; Whalen 1991, 1992; Pluymaekers et al 2005; Yun 2007).

In the current study, frequency matching was based on subjective judgments by 46 native speakers, none of whom participated in the production task. An objective frequency estimate is not possible for a number of reasons. Firstly, there is no electronic database of BHA, which is only a spoken vernacular. The situation in most Arabic-speaking communities is that dialects serve the best part of every-day oral interaction, while Standard Arabic is used for written transactions. Secondly, the Arabic writing system, where short vowels are not indicated, results in enormous ambiguity, if words are to be electronically extracted from a Standard Arabic database. For example, the orthographic form ذكر could stand for /ðikr/ ‘prayers’, /ðakar/ 'male’, /ðakar/ ‘mentioned’, /ðukir/ 'was mentioned', /ðakkar/ ‘reminded’, /ðukkir/ ‘was reminded’, /ðakkar/ ‘used the masculine form’, and /ðukkir/ ‘was used in the masculine form’.

Given this situation in BHA, only a subjective estimate of lexical frequency can be obtained. However, as a number of researchers suggest, subjective judgment can be as informative and reliable as an objective estimate of frequency (Carroll 1971; Gernsbacher 1984; Shapiro 1969; Snoeren et al 2006). Interestingly, in work where both are obtained, subjective judgment is found to be more strongly correlated with results of lexical decision tasks (Gordon 1985). Details of how subjective frequency has been estimated for the stimuli appear in Almihmadi (in preparation).

One obvious concern that remains is the morpho-syntacic status of the stimulus items. Specifically, words with $V2_{epenthetic}$ are predominantly nouns and those with $V2_{lexical}$ are almost all verbs. This follows from a general restriction in Arabic morphology where the canonical verb template is CVCVC, while nouns can be CVCC.[6]

The phonetics and phonology of individual words are known to be potentially sensitive to lexical category differences. For example, Smith (2001: 61) argues that "nouns show privileged phonological behaviour compared to words of other categories". Similarly, Conwell and Morgan (2007) report that English noun-vowels are significantly longer than the corresponding verb-vowels. Accordingly, the question that suggests itself for the current study is this: If epenthetics turn out to be acoustically different in a systematic way from the corresponding lexical vowels, how can we rule out the possibility that this difference is only a morpho-syntactic status effect?

It appears that there are at least two reasons that the morpho-syntactic status of a word in Arabic cannot be that influential. Firstly, word-order in Arabic is relatively free. Compare this with English, where nouns occur far more often clause-finally when they serve as an object of a verb phrase and clause-initially when serving as a subject of a verb phrase. Therefore, English nouns, unlike Arabic nouns or verbs, are naturally expected to be subject to edge lengthening (cf. Klatt 1975). Secondly, the syntactic structure of Arabic, particularly where nouns and verbs are involved, is such that many nouns and many verbs can stand on their own and form one-word sentences. This is in contrast to English, where most ambi-categorial words like 'run' would have to have different structures when it is a noun and when it is a verb. The noun version would have to occur following a determiner, an article, or a possessive pronoun. A one-word utterance 'run' is not grammatical on a noun-reading (Neil Smith, p.c. 2007) and if the verb is to occur on its own, it would normally have to be in the imperative. I return to the issue of morpho-syntactic status in the discussion section.

---

[6] There are a few verbs with a CVCC template, but there is no consensus as to whether or not C3 is a root radical or simply results from C2 gemination (see Gafos 2003; McCarthy 1986).

4.1.3 Procedures. Target items were orally elicited from each of the seven participants, using orally-presented questions that were designed for this purpose. Participants were instructed to say the response word followed by the word [tara] 'I think', as a two-word clause, with no intervening pause. The inclusion of [tara], which is both syntactically and semantically neutral, was to simplify segmentation before acoustic analysis. In a pilot, it was noticed that final sonorants and the preceding target vowel were so completely devoiced that any consistent, let alone, reliable segmentation would be very difficult to maintain. Since all target words end in a vowel followed by a sonorant, it seemed preferable to have a non-target word as pre-pausal.

All recordings were made using a Nagra ARES-M solid-state recorder and a Nagra NM-MICS-II premium quality clip-on mono microphone for ARES-M, placed a few inches away from the speaker's mouth. Recording was digitized at a sampling rate of 22.05KHz. Using Praat (Boersma & Weenink 2008), 588 tokens (14 pairs x 3 repetitions x 7 speakers) were acoustically analysed as follows. The acoustic analysis in this study targets V_2 in CV_1CV_2C test words. For the purposes of segmentation, V_2 is defined as the temporal interval between two boundary marks, B1 and B2. B1 is hand-inserted at the nearest zero crossing between the designated V_2 and a preceding [x], [ħ], [h], [b], [d], or [k]. B2 is hand-inserted between V_2 and a following [l], [r], [m], or [n]. Following Turk et al's (2006: 6) insightful recommendation, I have used "more zoomed out spectrogram displays" for locating a boundary region, and "more zoomed in waveform displays" for a more precise placement of the boundary mark within that region.

Generally, as in many acoustic studies (e.g., Alghamdi 1998, 2004; Ham 2001; Lavoie 2001; Ridouane 2007; Warner et al 2004), the main segmentation criterion in determining B1 and B2, respectively, was the onset and offset of the vowel F2. Occasionally, I had to consult F3 and its relative intensity (Skarnitzl 2009), in addition to F2 offset, for determining precisely where to insert B2 in V_2-liquid sequences.

Each labelled interval yielded acoustic data for five different parameters. These were mean F0 in Hz, mean intensity in dB, duration in ms, and F1 and F2 at midpoint. Extraction of the acoustic data was through a custom-written Praat script[7].

4.2 Results

As we can see from the figures in (7), mean and SD values of epenthetics and lexicals along each of the five acoustic parameters are very close, with exceedingly small differences overall.

(7) Mean and (SD) values of the acoustic parameters of the study for 'a' and 'i':

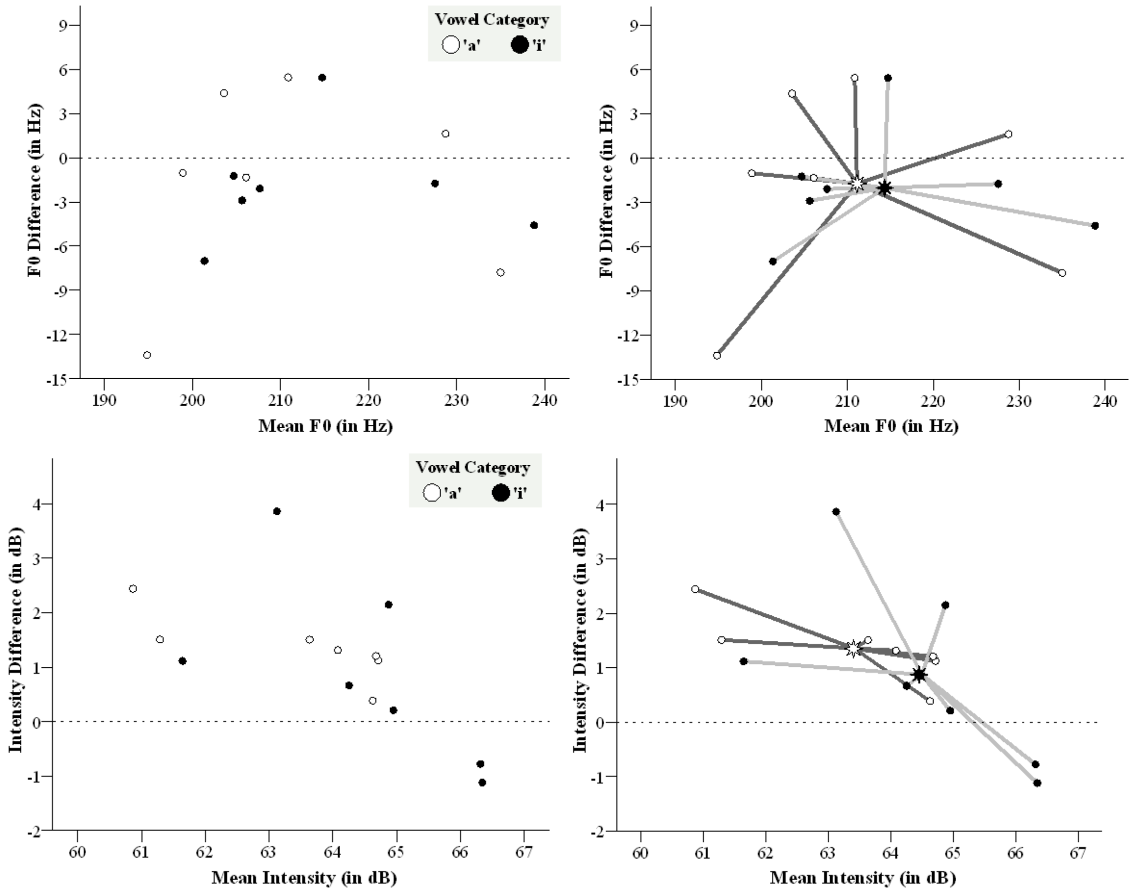
Parameter	[a]	/a/	[i]	/i/
F0 (Hz)	210 (16)	212 (15)	213 (14)	215 (14)
Intensity (dB)	64 (1.4)	62.7 (1.9)	64.9 (1.3)	64 (2.3)
Duration (ms)	80 (7.6)	78 (5)	71 (6)	69 (7.5)
F1 (Hz)	860 (48)	862 (38)	533 (46)	532 (40)
F2 (Hz)	1639 (160)	1636 (152)	2268 (266)	2265 (212)

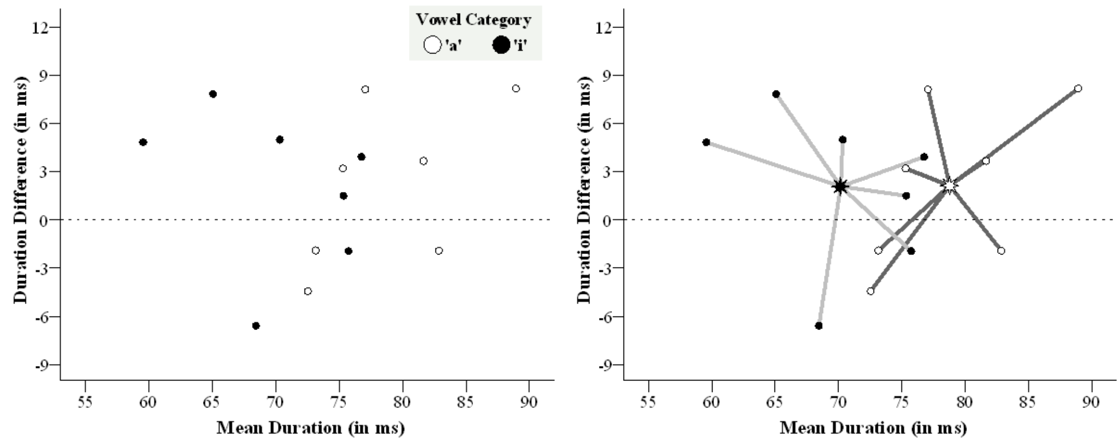
[7] The script was written by Yi Xu.

For both ‘a’ and ‘i’, the epenthetic vowel is longer in duration by 2ms, lower in pitch by 2Hz, and more intense with a larger mean paired difference for ‘a’ than for ‘i’. Mean paired differences along F1 and F2 range from 1–3Hz with F2 being 3Hz higher for epenthetics than for lexical vowels. I will describe the results of statistical significance tests below.

An important issue that we need to explore during initial examination of the data is the directionality issue. As described above, mean paired differences display the same directionality for both ‘a’ and ‘i’, along all the parameters except F1. However, we can learn a lot more than this by looking into Tukey’s mean-difference plots (e.g., Cleveland 1985; Tukey 1977). Tukey’s mean-difference plots for each of the acoustic parameters for ‘a’ and ‘i’ appear in (8) and (9). I have plotted mean paired difference ($\bar{x}_{PD}$) on the y-axis against the grand mean of each vowel category (($\bar{x}_{epe}$ + $\bar{x}_{lex}$)/2) for each individual speaker. For more clarity, I have reproduced these graphs and added to them group values and centroid spikes. Tukey’s plots offer a number of advantages for summarising paired datasets, the standard data type in phonetic research. For example, these plots show both the range of mean values and the range of mean paired differences for both individuals and groups. Moreover, the zero-difference reference line indicates how much and in what direction a paired difference departs from the no-difference zone.

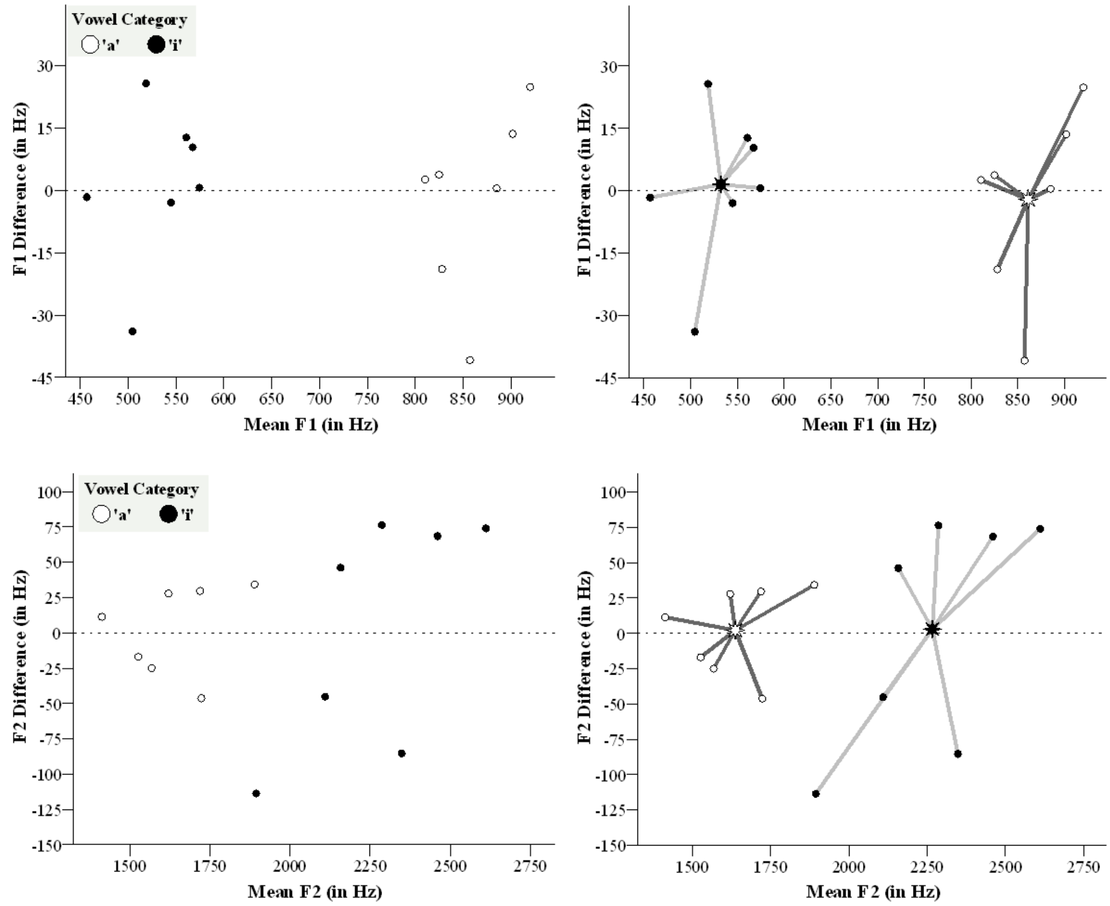
(8) Tukey’s mean-difference plots for F0, intensity, and duration for ‘a’ and ‘i’:
Difference = $\bar{x}_{epe} - \bar{x}_{lex}$; Mean = ($\bar{x}_{epe}$ + $\bar{x}_{lex}$)/2





(9) Tukey's mean-difference plots for F1 and F2 for 'a' and 'i':

Difference = $\bar{x}_{epe} - \bar{x}_{lex}$; Mean = $(\bar{x}_{epe} + \bar{x}_{lex})/2$



Looking at the graphs on the right-hand side in (8) and (9), we observe that the mean paired difference of the group data is closer to the zero-difference reference line than most individual paired differences. This seems to be the case for almost all the parameters for both 'a' and 'i'.

The graphs also show that individual $\bar{x}_{PD}$s go in opposite directions along all the parameters for both 'a' and 'i', except a-intensity, where all speakers produce more intense epenthetic [a]s than lexical /a/s. The pattern for i-intensity is very similar, with five speakers out of seven producing more intense [i]s than /i/s. A similar pattern obtains for i-duration, where five speakers produce longer [i]s than /i/s. However, we have the opposite pattern for i-F0, where six speakers have lower-pitched [i]s than /i/s.

Comparing these patterns and others to their corresponding group data, we can see how the distribution and magnitude of positive and negative individual paired differences define the group mean on the difference axis. That said, we should allow for the possibility that mean differences of similar magnitude but going in opposite directions can cancel each other out, thus levelling the group mean difference towards the zero-difference line, and increasing the variance of the data, at the same time. This is exactly what we observe in the case of i-F2, and to a lesser extent a-F2. The group F2 mean difference of [i]–/i/ is 3Hz, while the individual absolute mean differences are far larger than 3Hz, as given in (10) below.

(10)

Speaker	i-F2 $\bar{x}_{PD}$ (in Hz)
A	$|76|$
B	$|85|$
C	$|74|$
D	$|46|$
E	$|114|$
F	$|45|$
ALL	$|3|$

This cancelling effect can be demonstrated by a Wilcoxon Signed ranks test, which is the non-parametric equivalent of the paired t-test. Roughly speaking, the calculation procedure involves transforming mean differences into ranks, with the smallest difference assigned Rank 1, the next smallest Rank 2, and so on. These ranks inherit the sign (+ or -) of their untransformed values. Next, those ranks are summed and a Z-test statistic is computed, which can be used for null hypothesis testing involving the median paired difference. I give the summary rank statistics in (11) below. As can be seen from the table, adding up the positive and negative ranks for i-F2 amounts to zero.

(11) Wilcoxon Signed ranks test for i-F2

	'a'		'i'	
	$\sum - ranks$ /a/ < [a]	$\sum + ranks$ /a/ > [a]	$\sum - ranks$ /i/ < [i]	$\sum + ranks$ /i/ > [i]
F0	12	16	6	22
Intensity	28	0	20	8
Duration	20	8	20	8
F1	16	12	16	12
F2	16	12	14	14

Next, I ran separate paired t-tests on each parameter for each vowel category, correcting for multiple testing. The only statistically significant mean paired difference found was an intensity difference separating [a] from /a/: [mean difference=1.36dB; SD=.61dB; t(6)=5.87; p<.002]. All other mean paired differences for each vowel category were below statistical significance. All t-test outcomes appear in (12) below.

(12) Paired t-tests for 'a' and 'i'

	'a'		'i'	
	t(6)	p	t(6)	p
F0	-.68	.52	-1.4	.21
Intensity	5.87	.001*	1.3	.23
Duration	1.12	.31	1.12	.31
F1	-.25	.81	.22	.83
F2	.18	.86	.10	.93

*statistically significant at the Bonferroni-adjusted alpha level of 5%.

The results seem to paint a curious picture of the relation between the phonetics and phonology of neutralisation. Epenthetic [a] and lexical /a/, which behave the same in the phonology, are distinct in the phonetics. Conversely, epenthetic [i] and lexical /i/, which behave differently in the phonology, are phonetically identical. No less curious, though, is the apparent utilisation of a phonologically inactive acoustic cue for the purposes of preserving a contrast that seems to be completely neutralised in the phonology. I discuss these results next.

4.3 Discussion

In this section, I use the experimental data described above to (1) construct an account of the acoustics of vowel/zero neutralisation in BHA, and (2) discuss the various implications the results have for the phonetics-phonology relation.

For the purposes of this paper, neutralisation can be deemed phonetically complete if the five acoustic parameters studied here (F0, intensity, duration, F1, and F2) do not differ statistically significantly according to the underlying status of V2. With this in mind, we may conclude on the basis of the results above that the vowel/zero contrast is completely neutralised through [i]-epenthesis, but not through [a]-epenthesis. That is, [i] and /i/ are acoustically indistinguishable along all those parameters, while there is a statistically significant difference in intensity between [a] and /a/.

Abstracting away, for the moment, from the identity of the epenthesised vowel, we may conclude that vowel/zero neutralisation in BHA is acoustically both complete and incomplete. Put differently, the phonetic effect is variable between completeness and incompleteness. This finding has consequences for the genuineness question alluded to above. However, for space and scope limitations, I do not discuss these here. For more, see Almihmadi (in preparation).

The results for both [a]-epenthesis and [i]-epenthesis are not consistent with the predictions in (13) that the phonology of vowel/zero neutralisation seems to support.

(13) Assuming that the phonetics and phonology of neutralisation correlate closely, [a] and /a/, which behave phonologically the same, are predicted to be acoustically non-distinguishable, whereas [i] and /i/ are predicted to be acoustically different, since BHA phonology treats them differently.

What the results suggest is that neutralisation involving [a]-epenthesis, which is phonologically complete, is acoustically incomplete. Conversely, the distinction between epenthetic [i] and lexical /i/, which survives in the phonology, is lost in the acoustic signal at least along the parameters investigated here[8]. More generally, the phonetics and phonology of vowel/zero neutralisation in BHA do not mirror each other, for either vowel category.

Instead, they seem to dis-correlate, not just for one vowel category, but for both ‘a’ and ‘i’. The phonetics of neutralisation, as revealed by the statistical analysis of the experimental data, and the phonology of neutralisation, as documented by the pattern of interactions among phonological processes based on impressionistic data, do not agree on the completeness question. There is a statistically significant acoustic difference between [a] and /a/ that phonology overlooks. At the same time, phonology treats [i] and /i/ differently despite their acoustic non-distinguishablility.

Placing the acoustic completeness of neutralisation, as a null hypothesis, within the context of the phonetics-phonology relation highlights a potential conceptual problem with Null Hypothesis Significance Testing (henceforth NHST) procedures as currently applied in neutralisation laboratory studies. Rejecting or failing to reject the completeness null hypothesis for both ‘a’ and ‘i’, or rejecting it for ‘a’ while failing to reject it for ‘i’ will have this paradoxical implication for the phonetics-phonology relation: The phonetics and phonology of neutralisation both correlate and dis-correlate depending on the quality of the epenthetic vowel. That is, out of four logically possible scenarios for the phonetics-phonology relation, the NHST-analysed data present us with the weirdest. These scenarios appear in (14) below.

(14) Four logically possible scenarios for the phonetics-phonology relation according to neutralisation effects involving [a]-epenthesis and [i]-epenthesis

	‘a’	‘i’	Phonetics-Phonology Relation
Phonology	complete	incomplete	
Phonetics	complete	incomplete	close correlation
	complete	complete	partial correlation/non-correlation
	incomplete	incomplete	partial non-correlation/correlation
	incomplete	complete	complete non-correlation (anti-correlation)

Another interesting finding of the experiment involves the directionality issue, which has unduly been neglected in the phonetic research on neutralisation. A laboratory study of the phonetics of neutralisation can miss a lot by failing to take into consideration the directionality of the observed differences. Relying on a group averaged value, which is arithmetically calculated by dividing the sum of all individual data points by the number of data points used in the calculation, can result in an incomplete and potentially misleading generalisation. We have seen in the case of i-F2 that individual mean paired differences of similar magnitudes but going in opposing directions can cancel each other out in the calculation of the group arithmetic mean. This is reminiscent of the discrepancy reported in the literature between analyses based on group data and those based on data from each individual participant (see e.g., Dinnsen & Charles-Luce 1984; Gouskova & Hall 2009). What data should we use to draw inferences on the phonetics of neutralisation? Is this variability relevant? Is it lawful? Should we not doubt the tools of inference that we have

[8] It should be noted here that there can only be a limited set of acoustic parameters that can possibly be explored given the usual limitations on time and resources (cf. Kopkalli 1993; Jongman 2004). However, the fact that the very set of parameters investigated in this study has already displayed an epenthetic-lexical distinction for ‘a’ but not for ‘i’ seems to be interesting in and of itself, and worthy of further deliberation.

borrowed from psychology and the social sciences? Is the arithmetic mean the best measure of central tendency for phonetic data? I leave these questions open for the present. (See Almihmadi, in preparation).

Of relevance, though, is a puzzle that descriptive statistics presents regarding the direction of the acoustic differences between epenthetic and lexical vowels in terms of what is phonologically expected and what is phonetically observed. Phonologically, the "ideal epenthetic vowel is one that is least noticeable, i.e., one that is shortest and least sonorous", to quote Gouskova and Hall (2009: 219). Now the data we have of the phonetics of epenthetic and lexical vowels in BHA seem at odds with this common view in the phonological literature. Here, both epenthetic [a] and [i] are longer in mean duration and greater in mean intensity than the corresponding lexical vowels. Note that the argument still holds even if we choose to drop differences that have failed to reach statistical significance.

Does this take us back to the theme of dis-correlation between the phonetics and phonology of neutralisation? Not necessarily. Recall that the underlying status of the minimal pairs in the stimulus set coincides with a morpho-syntactic distinction. As I pointed out above, this follows form the morphology of BHA. A possible interpretation of the data in light of morpho-syntactic status is that noun-vowels in BHA are longer in duration and more intense than verb-vowels. Note that this pattern of results seems to be consistent with Conwell and Morgan's (2007) finding that noun-vowels in English are longer in duration than the corresponding verb-vowels. However, I should observe here that this conclusion is only true for descriptive statistics in my study, which display very negligible mean paired differences. The argument collapses as soon as descriptive statistics gives way to inferential statistics, which has only accorded statistical significance to the intensity difference between epenthetic [a] and lexical /a/. Why should only 'a' be longer in nouns than in verbs? Why shouldn't 'i' be like 'a' in this regard? Why should vowel quality matter for the morpho-syntax of any language? To salvage this morpho-syntactic account, we obviously need to find a morpho-syntactic argument for a quality-based distinction made in the phonetics of nouns and verbs. Returning to the realm of phonology and phonetics, we know that a quality-based distinction in vowel activity in phonetics and phonology is both demonstrable and real, as this paper has illustrated.

5 Conclusion

In this paper, I have presented a set of neutralisation data from a Bedouin Hijazi dialect of Arabic that seems to bring the phonetics and phonology of neutralisation face to face in a manner unreported before. Phonological and statistical analyses found both complete and incomplete neutralisation effects, from the point of view of both phonetics and phonology. However, phonologically incomplete neutralisation was found to be phonetically complete, whereas phonologically complete neutralisation was found to be phonetically incomplete. This apparent pattern of anti-correlation between the phonetics and phonology of neutralisation raises new concerns over how we approach, describe, and pass verdict on the phonetics of neutralisation. To advance our understanding of the phonetics of neutralisation, we need to address those concerns adequately in future research.

References

Alghamdi, M. (2004). *Analysis, synthesis and perception of voicing in Arabic*. Riyadh: Al-Toubah.
Alghamdi, M. (1998). A spectrographic analysis of Arabic vowels: A cross-dialect study, *Journal of King Saud University: Arts 10, 1*: 3-24.

Almihmadi, M. (in preparation). *Understanding the phonetics of neutralisation: A Variability-Field account of vowel/zero alternations in a Hijazi dialect of Arabic*.

Al-Mozainy, H. (1981). V*owel alternations in a Bedouin Hijazi Arabic dialect*: *Abstractness and stress*. PhD dissertation, University of Texas at Austin.

Al-Mozainy, H., R. Bley-Vorman, & J. McCarthy (1985). Stress shift and metrical structure. *Linguistic Inquiry 16*: 135-144.

Baković, E. (2007). A revised typology of opaque generalizations. *Phonology 24*: 217-259.

Barnes, J. (2006). *Strength and weakness at the interface: Positional neutralization at the interface*. Berlin: Mouton de Gruyter.

Baumann, M. (1995). The production of syllables in connected speech. PhD dissertation, University of Nijmegen.

Bermúdez-Otero, R. (1999). *Constraints interaction in language change: Quantity in English and German*. PhD dissertation, University of Manchester.

Blumstein , S . (1991). The relation between phonetics and phonology. *Phonetica 48*: 108–119.

Boersma, P. & D. Weenink (2008). *Praat: doing phonetics by Computer*. [Computer Programme: available at http://www.praat.org].

Brockhaus, W. (1995). *Final devoicing in the phonology of German*. (Linguistische Arbeiten 336). Tübingen: Niemeyer.

Butcher, A. R. (1995) The phonetics of neutrlisation: The case of Australian coronals. In J. Windsor Lewis (ed.), *Studies in general and English phonetics. Essays in honour of J.D. O'Connor*. London: Routledge. 10-38.

Carroll, J. (1971). Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior 10*: 722–729.

Clements, G. (1990). The role of the sonority cycle in core syllabification. In M. Beckman & J. Kingston (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge, Mass: CUP. 283-333

Cleveland, W. (1985). *The elements of graphing data*. New York: Chapman and Hall.

Colley, M. (2009). Investigating the incomplete neutralisation of flaps in North American English. *Journal of the Acoustical Society of America 126, 4*: 2181.

Conwell, E. & J. Morgan (2007). Resolving grammatical category ambiguity in Acquisition. In H. Caunt-Nulton, S. Kulatilake & I-h. Woo (eds.), *Proceedings of the 31st annual Boston University Conference on Language Development*. 117-128.

Dinnsen, D. (1985). A re-examination of phonological neutralization. *Journal of Linguistics 21*: 265–279.

Dinnsen, D. & J. Charles-Luce (1984). Phonological neutralization, phonetic implementation, and individual differences. *Journal of Phonetics 12*: 49-60.

Dmitrieva, O., A. Jongman & J. Sereno (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics 38, 3*: 483-492.

Ernestus, M. (in press). Categoricality and gradienece in phonology theory. In M. van Oostendorp, C. Ewen, B. Hume, and K. Rice (eds.), *Blackwell companion to phonology*. Oxford: Wiley-Blackwell.

Ernestus, M. & H. Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language 79*: 5-38.

Ettlinger, M. (2007). Incomplete neutralisation in Dutch opacity. *Journal of the Acoustical Society of America 122, 5*: 2997.

Farwaneh, S. (1995). *Directionality effects in Arabic dialect syllable structure*. PhD dissertation, University of Utah.

Fourakis, M. & G. Iverson (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica 41*: 140-149.

Fox, R. & D. Terbeek (1977). Dental flaps, vowel duration, and rule ordering in English. *Journal of Phonetics 5*: 27-34.

Gafos, A. (2003). Greenberg's asymmetry in Arabic: A consequence of stems in paradigms. *Language 79*: 317–355.

Gafos, A. & S. Benus (2006). Dynamics of Phonological Cognition. *Cognitive Science 30*: 1–39.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness and polysemy. *Journal of Experimental Psychology: General 113*: 256-281.

Goldrick, M. (2000). Turbid output representations and the unity of opacity. In: Masako Hirotani (ed.) *Proceedings of the North East Linguistics Society 30*. GLSA Publications, Amherst, MA. 231-246

Gouskova, M. & N. Hall (2009). Acoustics of epenthetic vowels in Lebanese Arabic. In S. Parker (ed.) *Phonological argumentation: Essays on evidence and motivation*. Equinox Publishing, London. 203-225.

Gordon, B. (1985). Subjective frequency and lexical decision latency function: Implications for mechanism of lexical access. *Journal of Memory and Language 24, 6*: 631-645.

Gurevich, N. (2004). *Lenition and contrast: the functional consequences of certain phonologically conditioned sound changes*. New York: Routledge.

Ham, W. (2001). *Phonetic and phonological aspects of geminate timing*. New York: Routledge.

Hansson, G. (2008). Effects of contrast recoverability on the typology of harmony systems. In P. Avery, B. E. Dresher & K. Rice (eds.), *Contrast in phonology: Perception and acquisition*. Berlin: Mouton de Gruyter. 115-141.

Jassem. W. & L. Richter (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics 17*: 317-325.

Jongman, A. (2004). Phonological and phonetic representations: The case of neutralization. In A. Agwuele, W. Warren & S-H. Park (eds.), *Proceedings of the 2003 Texas Linguistics Society Conference*. Somerville, MA: Cascadilla Proceedings Project. 9-16

Kiparsky, P. (1973). Abstractness, opacity, and global rules. In O. Fujimura (ed.), *Three dimensions in linguistic theory*. Tokyo: TEC. 57–86

Kiparsky, P. (1971). Historical Linguistics. In W. O. Dingwall (ed.), *A survey of linguistic science*. College Park: University of Maryland Linguistics Program. 576-642

Klatt, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics 3*: 129-140.

Kopkalli, H. (1993). *A phonetic and phonological analysis of final devoicing in Turkish*. PhD dissertation, The University of Michigan.

Lavoie, L. (2001). *Consonant strength. phonological patterns and phonetic manifestations*. New York: Garland.

Manaster Ramer, A. (1996). A letter from an incompletely neutral phonologist. *Journal of Phonetics 24*: 477–489.

McCarthy, J. (2007). *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.

McCarthy, J. (1999). Sympathy and phonological opacity. *Phonology 16*: 331–399.

McCarthy, J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry 7*: 207-263.

Munson, B. & N. Solomon (2004). The effect of phonological neighbourhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research 47*: 1048-1058.

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In D. Docherty & D. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: CUP. 261-280.

Nolan, F. (1986). The implications of partial assimilation and incomplete neutralisation. *Cambridge Papers in Phonetics and Experimental Linguistics 5*: N1-N11.

Nycz, J. (2005). The dynamics of near merger in accommodation. *Proceedings of ConSOLE XIII, 2005*: 1-13.

O'Dell, M. & R. Port (1983). Discrimination of word-final voicing in German. *Journal of the Acoustical Society of America 73*: supplement 1.

Oh, E. (1998). Stress patterns of Bedouin Hijazi Arabic: An OT account. *Kansas Working Papers in Linguistics 23*: 17-26. University of Kansas. Linguistics Graduate Student Association.

van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua 118, 9*: 1362-1374.

Pluymaekers, M., M. Ernestu & H. Baayen (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America 118, 4*: 2561-2569.

Port, R. & P. Crawford (1989). Pragmatic effects on neutralization rules. *Journal of Phonetics 16*: 257-282.

Ridouane, R. (2007). Gemination in Tashlhiyt Berber: An acoustic and articulatory study. *Journal of the International Phonetic Association 37, 2*: 119-142..

Shapiro, B. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behaviour 8*: 248–251.

Sheldon, D. (1973). A short experimental investigation of the phonological view of the writer-rider contrast in U.S. English. *Journal of Phonetics 1*: 339–346.

Skarnitzl, R. (2009). Challenges in degmenting the Czech lateral liquid. In A. Esposito & R. Vich (eds.) *Cross-Modal Analysis of Specch, Gestures, Gaze and facial Expressions. Lecture Notes in Computer Science 5641*: 162: 172.

Slowiaczek, L., & D. Dinnsen (1985). On the neutralizing status of Polish word final devoicing. *Journal of Phonetics 13*: 325-341.

Smith, J. (2001). Lexical category and phonological contrast. In R. Kirchner, J. Pater, & W. Wikely (eds.), *PETL 6: Proceedings of the Workshop on the Lexicon in Phonetics and Phonology*. Edmonton: University of Alberta. 61-72.

Snoeren, N., P. Hallé & J. Segui (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics 34*: 241–268.

Tukey, J. (1977). *Exploring data analysis*. Addison-Wesley.

Turk, A., S. Nakai, M. Sugahara (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schlieβer (eds.), *Methods in empirical prosody research*. Berlin: Walter de Gruyter. 1: 27.

Warner, N., E. Good, A. Jongman & J. Sereno (2006). Orthographic vs. morphological incomplete neutralization effects. Letter to the editor, *Journal of Phonetics 34*: 285-293.

Warner, N., A. Jongman, J. Sereno & R. Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics 32*: 251-276.

Whalen, D. (1992). Further results on the duration of infrequent and frequent words. *Journal of the Acoustical Society of America 91, 4:* 2339–2340.

Whalen, D. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America 90, 4*: 2311.

Yu, A. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology 24*: 178-214.

Yun, G. (2007). Word frequency, stress and coarticulation in English. *Studies in Phonetics, Phonology and Morphology 13*, *2*: 315-331.

Zue, V. & M. Laferriere (1979). Acoustic study of medial t, d in American English. *Journal of the Acoustical Society of America 66*: 1039–1050.