

Learning to detect a tone in unpredictable noise

Pete R. Jones and David R. Moore

*MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD,
United Kingdom
p.r.jones@ucl.ac.uk, david.moore2@cchmc.org*

Daniel E. Shub

*School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD,
United Kingdom
daniel.shub@nottingham.ac.uk*

Sygal Amitay

*MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD,
United Kingdom
sygal@jhr.mrc.ac.uk*

Abstract: Eight normal-hearing listeners practiced a tone-detection task in which a 1-kHz target was masked by a spectrally unpredictable multitone complex. Consistent learning was observed, with mean masking decreasing by 6.4 dB over five sessions (4500 trials). Reverse-correlation was used to estimate how listeners weighted each spectral region. Weight-vectors approximated the ideal more closely after practice, indicating that listeners were learning to attend selectively to the task relevant information. Once changes in weights were accounted for, no changes in internal noise (psychometric slope) were observed. It is concluded that this task elicits robust learning, which can be understood primarily as improved selective attention.

© 2014 Acoustical Society of America

PACS numbers: 43.66.Ba, 43.66.Dc [QJF]

Date Received: November 21, 2013 Date Accepted: January 27, 2014

1. Introduction

Detection thresholds for a fixed-frequency sinusoid deteriorate by 20–50 dB in the presence of a spectrally unpredictable multitone complex (Kidd *et al.*, 2007). Such masking cannot be explained by overlapping activity in peripheral auditory filters because it occurs even when the masker is energetically weak or spectrally distal (“across-channel interference”). Instead it appears driven by higher order factors, such as the degree of masker uncertainty (Neff and Callaghan, 1988) and the similarity between target and masker (Lee and Richards, 2011).

Whether listeners can learn to reduce such “informational” masking is of general interest, particularly because hearing in unpredictable environments is a prominent source of difficulty and dissatisfaction for hearing-impaired listeners (Gatehouse and Noble, 2004). However, learning on unpredictable masking tasks has tended to be obscured by the use of highly experienced participants. Two studies by Neff and colleagues *did* explicitly examine practice effects (Neff and Callaghan, 1988; Neff and Dethlefs, 1995). These indicated that learning occurs in a minority of listeners but that performance generally remains “remarkably stable” across sessions. Notably though, listeners were not completely naive to the task. The listeners in the later study by Neff and Dethlefs (1995) completed 600 practice trials prior to testing, while those in the earlier study by Neff and Callaghan (1988) had extensive experience (>10 h) of related masking tasks. Moreover, listeners generally completed fewer than 1000 test trials. Given that auditory learning tends to be greatest early in training (Hawkey *et al.*,

2004), and may extend over many thousands of trials, the full extent of learning therefore remains uncertain.

The mechanisms subserving learning also remain unclear, and understanding them is important for the design of effective training schedules. One possibility is that levels of unpredictable masking are determined by the width of the listener's "window of attention" (i.e., the spectral range over which auditory filter activity is integrated; Lutfi, 1993). Learning may therefore represent a narrowing of this window with listeners learning to give weight only to the target region. Alternatively, learning may be the result of reduced internal noise (i.e., reduced variability in the listener's decision variable) as has been suggested previously for other perceptual-judgment tasks (Jones *et al.*, 2013). We evaluated both of these possibilities in the present study, using a two-step procedure (cf. Berg, 2004). First, reverse correlation was used to estimate the relative weight the listener gave to each spectral region. The observed weights were compared to the ideal to derive a measure of efficiency. Second, the estimated weights were used to derive the listener's trial-by-trial decision variable, DV , and a psychometric curve was fitted to the probability of responding "interval 2" as a function of DV . Because the goodness of the weighting strategy was partialled out, the slope parameter could be interpreted as an index of internal noise.

2. Methods

2.1 Listeners

Eight listeners (five female) participated. They were aged 19–26 yr and had no prior experience in psychophysical tasks. All had audiometrically normal hearing [≤ 20 dB hearing level (HL) bilaterally, at 0.25-8-kHz octaves]. Participants were recruited through advertisements placed around the Nottingham University campus, and received £7.5/h.

2.2 Task and procedure

The task was two-alternative, forced-choice (2AFC) tone detection in which participants were asked to "pick the interval containing the target tone." Each trial consisted of two 300-ms observation intervals, separated by a 500-ms interstimulus interval. Listeners had an unlimited time to respond, and responses were followed by 250 ms of visual feedback (a "happy" or "sad" smiley face).

In each block, a two-down one-up adaptive track was used to derive an estimate of the listener's 70.7% correct detection threshold, either in noise or in quiet. The level of the target tone was initialized at 60 dB sound pressure level (SPL). It was adapted in steps of 8 dB until the second reversal and 2 dB thereafter. Each block consisted of 50 trials. The number of trials was fixed rather than the number of reversals to ensure that all listeners received the same amount of practice. Before each block listeners were presented with the target in quiet as a reminder.

Each session lasted approximately 45 min and consisted of 16 noise blocks and 2 quiet blocks. The 18 blocks were presented in random order with a rest break after the 10th block. Listeners completed five sessions over five consecutive days (4500 trials). Initially, participants also completed one practice trial in quiet and three practice trials in noise. To highlight the task demands, during this practice the stimuli durations were increased to 800 ms, and noises were attenuated to 50 dB SPL.

2.3 Stimuli and apparatus

The target was always a 1-kHz sinusoid, randomly assigned with equal probability to one of two observation intervals. In noise blocks, 30-component multitone complexes were also presented in each interval. All stimuli were 300 ms in duration, including 10-ms \cos^2 on/off ramps, and were presented diotically over Sennheiser HD 25-I headphones.

The frequency, phase, and amplitude of the noise components were independently randomized on every presentation. Phases were randomly drawn from a rectangular distribution. Amplitudes were randomly drawn from a Rayleigh distribution and

normalized so that the total masker level was always 60 dB SPL. Frequencies were randomly drawn without replacement from 715 candidates, log-distributed between 223 and 4490 Hz, excluding a third-octave notch geometrically centered on the target frequency (891–1120 Hz). This notch served to minimize energetic masking, and was slightly greater than the average equivalent rectangular band [ERB] at 1 kHz (Glasberg and Moore, 1990).

Stimuli were digitally synthesized in MATLAB v7.4 (2007a, The MathWorks, Natick, MA) using 44.1-kHz sampling and 24-bit quantization. Digital-to-analog conversion was performed by a PCI sound card (Darla Echo; Echo Digital Audio, Carpinteria, CA). Listeners responded via a button box and were tested individually in a double-walled sound-attenuating booth.

2.4 Measures

Detection thresholds were calculated independently for each track as the mean target level (dB) at the last four reversals. With noise blocks, the amount of masking was computed by subtracting the mean detection threshold in quiet (averaged over all blocks).

Relative weights were calculated via multiple logistic regressions, as per Alexander and Lutfi (2004). The dependent variable was the listener's binary response ("interval 1" or "interval 2"). The independent variables were the differences in decibel level between the corresponding spectral region in each observation interval. In instances where there was no energy in a spectral region, the level was set to 0 dB. The weights, ω , were the regression coefficients, normalized so that their magnitudes summed to 1. As in Alexander and Lutfi (2004), the efficiency of the weight strategy was calculated as one minus the root mean square (RMS) difference between observed and ideal weights. The ideal strategy was to assign a weight of unity to the target bin and zero-weight elsewhere, which would yield an efficiency value of one.

The magnitude of the internal noise, σ_{int} , was calculated as the standard deviation of a zero-mean cumulative normal distribution, fitted to the binned probability of a listener responding interval 2 as a function of the estimated decision variable, DV . The DV was defined as $\sum_{i=1}^n \omega_i \Delta L_i$, where ΔL_i represents the energetic level difference (in dB) in the i th spectral bin, and ω_i is the corresponding relative weight coefficient. Psychometric fits were made using PSIGNIFIT v2.5.6: A MATLAB toolbox that implements the maximum-likelihood method described by Wichmann and Hill (2001).

3. Results and discussion

3.1 Learning

Detection thresholds in quiet and in noise are plotted for individuals in Fig. 1. In quiet, no learning was observed with all individuals well-described by linear regressions

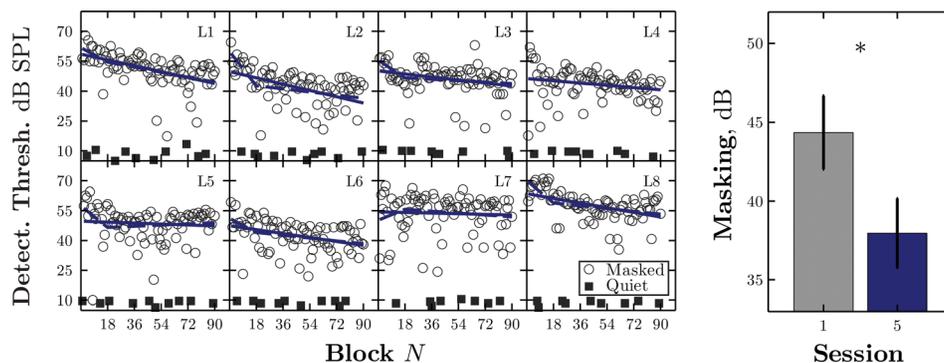


Fig. 1. (Color online) (Left) 70.7% correct detection thresholds as a function of block measured in quiet (filled squares) or in noise (open circles). Solid lines represent least square linear fits to the noise data. Dashed lines denote broken-stick fits, inflected after session 1. (Right) Group-mean (± 1 SE) change in masking.

with near-zero slope [$\beta_{\mu}=0.01$; all $p > 0.05$]. In contrast, substantial learning was observed in the masked condition. Linear fits yielded significant negative slopes for all but two (L5, L7) listeners [$p < 0.05$] with improvement rates ranging from -0.02 to -0.18 dB/block. Some data were better fit by broken-stick functions, suggesting a short initial phase of rapid learning followed by a protracted period of more gradual learning. However, the improvements in fit were small [$\Delta r^2/\Delta_{df} < 1$] for all but L5, indicating that learning may be more gradual than in more basic auditory tasks, such as frequency discrimination (Hawkey *et al.*, 2004). Grand mean masking decreased from 44.3 dB in session one to 38.0 dB in session five [$t(7) = 4.09$, $p = 0.005$].

We conclude that the ability to detect a tone in unpredictable noise improves robustly with practice. Furthermore because the performance of some listeners (e.g., L1 and L4) had not plateaued by the end of the study, further learning may have been possible. This release from across-channel interference is consistent with Buss (2008), who found that six of eight listeners improved with practice on an analogous intensity-discrimination in unpredictable-noise task. However, widespread learning appears contrary to Neff and Dethlefs (1995), where tone-in-unpredictable-noise detection-thresholds did not improve in most listeners. This is likely because those listeners completed 600 trials practice prior to training. Accordingly, when the first 600 trials were excluded in the present study, regression slopes were significantly shallower [$t(7) = -2.43$, $p = 0.046$] and failed to differ from zero (no-learning) in five of eight listeners [five $p > 0.05$]. The present results are therefore in good agreement with those of Neff and Dethlefs (1995).

There was substantial individual variability in masking: 36.6–58.0 dB in session one, 31.6–50.0 dB in session five. Previous authors have wondered whether such individual differences are reduced by training (e.g., Durlach *et al.*, 2003). As in Neff and Callaghan (1988), we found little evidence of that here; between-subject variation in masking was approximately constant across all sessions with the greatest variability occurring in session four where group mean masking was actually lowest. There was some indication that variability within listeners (intra-session masking S.D.) decreased with practice, but this change was not significant [$t(7) = 2.27$; $p = 0.058$, n.s.].

3.2 Mechanisms of learning

Estimates of weights are shown for individuals in Fig. 2. Most listeners gave greatest weight to the target bin (1 kHz), but listeners also gave substantial weight to irrelevant spectral information. Overweighting was particularly prevalent at the upper fringe of the stimulus, consistent with previous reports that the higher-frequency components of a complex tend to be resolved more accurately (Watson *et al.*, 1975). A repeated

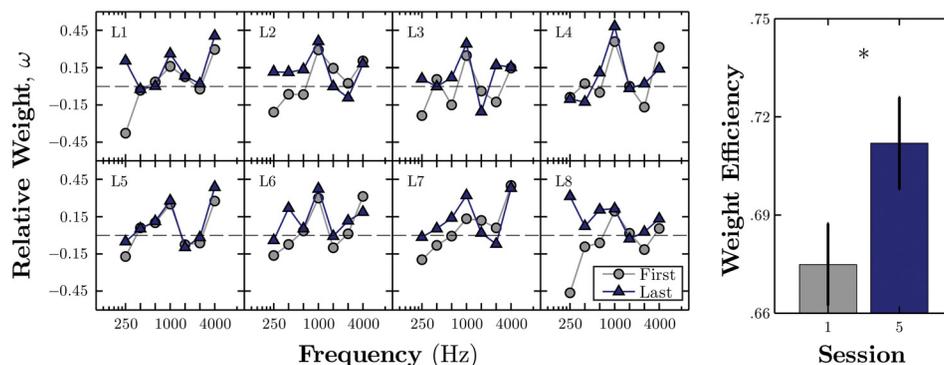


Fig. 2. (Color online) (Left) Individual weights for the first (circles) and last (triangles) session. Each point represents the geometric center of a third-octave spectral bin. The target signal was always a 1-kHz sinusoid, so the optimal strategy was to always give unit weight to the 1-kHz bin, and zero weight elsewhere. (Right) Group-mean (± 1 SE) change in weight efficiency.

measures analysis of variance (rm-ANOVA) yielded a significant main effect of session on weight efficiency [$F(4, 28) = 3.48$, $p = 0.020$], indicating that listeners' weighting strategies improved with practice. Thus after practice listeners predicated their decisions more selectively on the target spectral region. However, there was no straightforward relationship between the microstructure of the weights and learning. Thus L1 most overweighted high frequencies after training but exhibited the second most learning, whereas L5 and L7 had a similar high-frequency bias but exhibited the shallowest learning slopes (cf. Fig. 1). It may therefore be that individuals learn distinct listening strategies even on relatively simple perceptual-judgment tasks.

To evaluate changes in internal noise, cumulative Gaussians were fitted to listeners' responses as a function of the estimated trial-by-trial DV . Group-mean estimates of σ_{int} did not systematically vary across session [$F(4, 28) = 0.20$, $p = 0.937$, n.s.], indicating that internal noise magnitude was not diminished by practice. However, as shown in Fig. 3, there was variability between listeners. For example, listeners L1 and L4 exhibited a marked decrease in internal noise, as indicated by steeper psychometric slopes in session five. Conversely, listener L8 shows very little change (but did exhibit substantial learning; cf. Fig. 1).

Alternate psychometric fits were also made assuming ideal weights throughout (i.e., as a function of target level). These models did indicate a reduction in internal noise [$F(4, 28) = 5.49$, $p = 0.002$, $\eta_p^2 = 0.44$]. However they gave a markedly poorer account of the raw data with significantly greater deviance between model and data [rm-ANOVA; $F(1, 63) = 11.02$, $p = 0.013$, $\eta_p^2 = 0.61$]. Thus we favor the combined weight+internal noise account here, both as it provides empirically stronger fits and because of its potential to predict how performance will change depending on how spectral composition of the stimulus varies.

In summary, training produced significant improvements in weight efficiency but not internal noise—listeners learned with practice to attend selectively to the task-relevant information. Finally, to examine whether these changes differed in *effect* as well as significance, Monte Carlo simulations were run using the observed group-mean changes in weight efficiency and internal noise. These simulations followed the same test procedure as human listeners (i.e., same n trials, n listeners, etc.), and the simulations similarly responded to the greatest weighted-sum-energy. Thus the decision rule was to respond interval 1 only when $[\sum_{i=1}^n \omega_i \Delta L_i + \sigma_{int}/\sqrt{2}] > 0$.

When internal noise was held constant at its mean session 1 value, and weight efficiency was varied from its session 1 to session 5 value, mean-masking decreased by an amount similar to that observed empirically [*Similar*: -7.2 dB; *Empirical*: -6.7 dB]. Conversely, varying internal noise and holding weight efficiency constant produced no

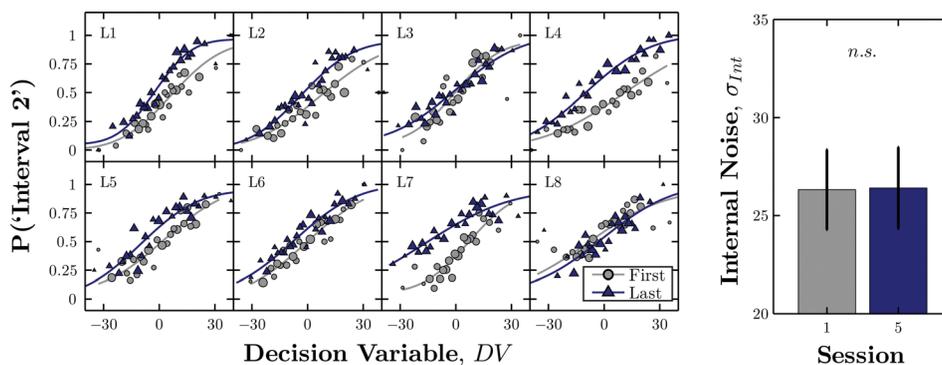


Fig. 3. (Color online) (Left) Individual psychometric fits for the first (circles) and last (triangles) session, relating probability of responding “interval 2” to the estimated trial-by-trial decision variable. The standard deviation parameter, σ , of each cumulative Gaussian fit was taken as a measure of internal noise magnitude. (Right) Group-mean (± 1 SE) change in internal noise magnitude.

significant change in masking [$p = 0.399$]. These results suggest that improvements in selective attention primarily drive learning on this task.

4. Conclusions

- (1) Masking by unpredictable noise is reduced by practice in most listeners. Much of this learning occurs rapidly, within the first 600 trials, but learning continues in some listeners for several thousand trials.
- (2) Selective attention underlies improvement on this task, with listeners learning to appropriately weight the task-relevant information. No significant changes in internal noise were observed once decision-weights were taken into account.

Acknowledgment

This work was supported by the Medical Research Council, UK (Grant No. U135097130).

References and links

- Alexander, J. M., and Lutfi, R. A. (2004). "Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights," *J. Acoust. Soc. Am.* **116**, 2234–2247.
- Berg, B. G. (2004). "A molecular description of profile analysis: Decision weights and internal noise," *J. Acoust. Soc. Am.* **115**(2), 822–829.
- Buss, E. (2008). "Across-channel interference in intensity discrimination: The role of practice and listening strategy," *J. Acoust. Soc. Am.* **123**(1), 265–272.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). "Note on informational masking (L)," *J. Acoust. Soc. Am.* **113**(6), 2984–2987.
- Gatehouse, S., and Noble, W. (2004). "The speech, spatial and qualities of hearing scale (SSQ)," *Int. J. Audiol.* **43**(2), 85–99.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**(1–2), 103–138.
- Hawkey, D. J. C., Amitay, S., and Moore, D. R. (2004). "Early and rapid perceptual learning," *Nat. Neurosci.* **7**(10), 1055–1056.
- Jones, P. R., Shub, D. E., Moore, D. R., and Amitay, S. (2013). "Reduction of internal noise in auditory perceptual learning," *J. Acoust. Soc. Am.* **133**(2), 970–981.
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2007). "Informational masking," in *Auditory Perception of Sound Sources*, edited by S. A. S. Christophe Micheyl and A. J. Oxenham (Springer Science+Business Media, New York), pp. 143–189.
- Lee, T. Y., and Richards, V. M. (2011). "Evaluation of similarity effects in informational masking," *J. Acoust. Soc. Am.* **129**(6), EL280–EL285.
- Lutfi, R. A. (1993). "A model of auditory pattern analysis based on component-relative-entropy," *J. Acoust. Soc. Am.* **94**(2), 748–758.
- Neff, D. L., and Callaghan, B. P. (1988). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," *J. Acoust. Soc. Am.* **83**(5), 1833–1838.
- Neff, D. L., and Dethlefs, T. M. (1995). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," *J. Acoust. Soc. Am.* **98**(1), 125–134.
- Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbassat, C. A. (1975). "Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals," *J. Acoust. Soc. Am.* **57**(5), 1175–1185.
- Wichmann, F. A., and Hill, N. J. (2001). "The psychometric function: I. fitting, sampling, and goodness of fit," *Atten. Percept. Psychophys.* **63**(8), 1293–1313.