Effect of stimulus bandwidth and duration on monaural envelope correlation perception

Emily Buss^{a)}

*Department of Otolaryngology/Head and Neck Surgery,
University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599
ebuss@med.unc.edu*

Huanping Dai

*Department of Speech, Language, and Hearing Sciences,
University of Arizona, Tucson, Arizona 85721
hdai@email.arizona.edu*

Joseph W. Hall III

*Department of Otolaryngology/Head and Neck Surgery,
University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599
jwh@med.unc.edu*

Abstract: Monaural envelope correlation perception is the ability to discriminate between stimuli composed of two or more bands of noise based on envelope correlation. Sensitivity decreases as stimulus bandwidth is reduced below 100 Hz. The present study manipulated stimulus bandwidth (25–100 Hz) and duration (25–800 ms) to evaluate whether performance of highly trained listeners is limited by the number of inherent modulation periods in each presentation. Stimuli were two bands of noise, separated by a 500-Hz gap centered on 2250 Hz. Performance improved reliably with increasing numbers of envelope modulation periods, although there were substantial individual differences.

© 2014 Acoustical Society of America

[QJF]

Date Received: July 18, 2014 **Date Accepted:** November 20, 2014

1. Introduction

Monaural envelope correlation perception (MECP) is the ability to discriminate between stimuli based on the correlation of temporal envelopes between two or more bands of noise that are separated in frequency (Richards, 1987). Sensitivity to envelope correlation and the subjective cue used to discriminate correlated vs random bands is reported to be similar whether two bands are presented monaurally or to opposite ears, indicating that MECP probably does not depend on beating between bands (Richards, 1988, 1989; Buss *et al.*, 2013). Recently Buss *et al.* (2013) demonstrated above-chance MECP for pairs of bands ranging from 25 to 1600 Hz wide; in all conditions, the two bands were separated by a 500-Hz gap centered on 2250 Hz. The best performance was seen for bandwidths of 100–400 Hz, with reduced sensitivity for both narrower and wider bandwidths. Buss *et al.* (2013) discussed several factors that could explain the results of MECP for wide bands (>400 Hz), including a possible role of peripheral auditory filtering and combination of information across filters. The present report focuses on MECP for narrow bands (≤ 100 Hz), for which each band falls within a single auditory filter.

Moore and Emmerich (1990) proposed that decreasing stimulus bandwidth below 100 Hz in the MECP paradigm could degrade performance by reducing the

^{a)}Author to whom correspondence should be addressed.

number of periods of inherent modulation available to the listener. That study measured MECF for two bandwidths (25 and 100 Hz) and two durations (100 and 500 ms). Performance was best for the wide bandwidth and long duration, worst for the narrow bandwidth and short duration, and intermediate for the two other combinations of duration and bandwidth. Whereas increasing duration improved d' by a factor of $\sim\sqrt{n}$, the effect of increasing bandwidth fell short of the predicted *square root of n* improvement. Moore and Emmerich suggested that the more modest effect of bandwidth seen in their data could reflect a reduced ability to utilize the higher rate envelope fluctuations associated with the wider bandwidth (Viemeister, 1979).

One goal of the present experiment was to systematically evaluate the effects of bandwidth and duration on MECF for sub-critical bands of noise. Since the envelope modulation rate of noise is proportional to its bandwidth, the prediction was that changing either the bandwidth or the duration by a factor of 2 should have similar effects on performance, with the caveat that the effect of increasing bandwidth may be somewhat limited by the ability to make use of rapid envelope fluctuations. A secondary goal of this experiment was to evaluate the consistency of listener responses to particular stimulus samples. If listener responses to frozen noise are consistent, then it may be possible to characterize the stimulus features, and possibly perceptual cues, associated with easy vs difficult samples. In particular, for a fixed duration and bandwidth, random samples of noise can have different numbers of inherent modulation periods. If the number of modulation periods is an important stimulus feature for MECF, then samples with more modulation periods should be associated with better listener performance.

2. Methods

Stimuli were pairs of band-pass noise samples, each either 25-, 50-, or 100-Hz wide. The low band had a fixed upper edge of 2000 Hz, and the high band had a fixed lower edge of 2500 Hz, such that there was a 500-Hz gap between bands regardless of bandwidth. Fully correlated bands were generated in the frequency domain, with the same arrays of random draws defining the real and imaginary components for corresponding bins in the low- and high-frequency bands. There were 100 sets of low- and high-frequency bands with correlated envelopes at each bandwidth, each 820 ms in duration. Stimuli presented to the listener were the sum of a low and a high band, drawn either from a single set of bands (with fully correlated envelopes) or from different sets (with random envelopes). Both bands were gated on at the beginning of the frozen samples and gated off after the specified duration using 10-ms raised-cosine ramps. Stimulus duration, measured from the 6-dB-down points, was 25, 50, 100, 200, 400, or 800 ms. The overall level was 65 dB sound pressure level (SPL). In addition to the low and high bands, there was a continuous low-pass noise to mask potential low-frequency distortion cues (Buss *et al.*, 2013). The low-pass noise was generated with a 6th order Butterworth filter and a 1000-Hz cutoff, and it was presented at 65 dB SPL. All stimuli were generated in MATLAB, loaded into a real-time processor (RP2, TDT), routed to a headphone buffer (HB7, TDT), and presented to the listener's left ear using circumaural headphones (Sennheiser, HD 265 linear).

The task was a two-alternative forced choice. One interval, selected at random, contained a pair of bands with fully correlated envelopes, and the other contained randomly paired bands, for which the correlation could fall anywhere between -1 and 1 . The correlated and random stimuli presented on each trial were randomly selected without replacement. The listener's task was to select the interval containing the correlated bands. Onsets associated with the two intervals were separated by 1.2 s, such that the inter-stimulus interval decreased with increasing stimulus duration. Lights on a hand-held response box indicated listening intervals and provided correct-answer feedback immediately following each listener response. Stimulus bandwidth and duration were held constant within a block of 50 trials. Each listener completed five such blocks (a total of 250 trials) in each condition. Data collection began with

the widest (100-Hz) bandwidth, and progressed to narrower bandwidths. Within bandwidths, testing began with the longest (800-ms) duration, and progressed to shorter durations. Testing for a particular bandwidth was discontinued if listener performance fell at or below 55% correct. While this easy-to-hard test order could result in practice effects, this possibility seems unlikely given that all listeners had extensive prior experience with this task.

After completing the first phase of data collection, additional data were collected for the 100-Hz stimulus in order to assess listeners' responses to particular frozen noise samples. The stimulus duration used for each listener was associated with ~75% correct in the first phase of data collection. The rationale for using individualized durations was to ensure that listeners were neither at floor nor ceiling, such that reliable differences in listener responses to particular stimuli could be observed. This extended data collection used just the first 25 sets of frozen noise stimuli. In contrast to the data collected up to this point, the association between correlated and random bands within a trial was itself frozen. That is, each of the 25 correlated-envelope stimuli was associated with a single random-envelope stimulus; these stimuli were always presented together within a trial, although the order of presentation within the trial was random. Eliminating stimulus variability in this way was considered desirable in the context of assessing the reliability and consistency of listener responses. Each block of 50 trials included two presentations of each of the 25 pairs of correlated and random stimuli, and listeners completed 10–20 blocks, depending on availability.

Five normal-hearing listeners were recruited, with thresholds in quiet of 20 dB hearing level or better for octave frequencies between 250 and 8000 Hz (ANSI, 2010) in the test ear. Their ages ranged from 35 to 63 years. All listeners had more than 200 h of psychoacoustic listening experience, including at least 15 h of exposure to MECF stimuli. Listeners L1, L2, and L5 had previously participated in the MECF study of Buss *et al.* (2013) (identified in that report as L4, L2, and L5, respectively).

3. Results

The left column of Fig. 1 shows percent correct plotted as a function of stimulus duration for each listener, with the more sensitive listeners at the top of the figure. Symbols indicate the stimulus bandwidth, as defined in the legend. Listeners L4 and L5 were highly variable and performed relatively poorly, failing to reach 95% correct in any condition, despite 20 h or more of practice on various MECF conditions. The data of listeners L2 and L3 were broadly similar to the expectation that performance should improve with increasing number of modulation periods, either via increased bandwidth or increased duration: for a given bandwidth, performance improved with increasing duration, and performance tended to be better for the wider bandwidths. The results of L1 are different from those of the other listeners. For this listener, effects of bandwidth and duration were more modest. At the short durations, where L2 and L3 performed at chance, L1 scored above 70% correct for all three bandwidths. Comparing listeners at the duration associated with ~80% correct for the 100-Hz bandwidth, scores for L1 were limited to a range of just 7 percentage points across the three bandwidths, in contrast to the wider range of scores in the data of L2 and L3 (18 and 28 percentage points, respectively).

Comparisons between the present results and those of Buss *et al.* (2013) are complicated somewhat by the fact that the previous study used a three-alternative forced choice, whereas the present study used a two-alternative forced choice. However, comparisons with previous data support the observation that L5 is less sensitive to MECF than L1 and L2. For example, in the 100-Hz bandwidth conditions with the low-pass noise in the previous experiment, scores for L1 and L2 differed by less than one percentage point, whereas the score for L5 was 10 percentage points lower. Data for L1 and L2 were comparable in the previous experiment.

The joint effects of bandwidth and duration were evaluated via data fits. For each listener, the percent correct (PC) scores were expressed as a function of WD , the

Fig. 1. (Color online) Percent correct is plotted for each listener, with symbol shape indicating stimulus bandwidth (as defined in the legend). The left column shows results as a function of stimulus duration, and the right column shows results as a function of $0.64WD$, the expected number of envelope modulation periods per stimulus. Individual listeners' results are plotted in separate rows. Data were fitted with a multiple-look model, wherein d' improves with $\sqrt{n}$ and n is the expected number of periods of inherent modulation. Thick lines in the right column indicate data fits, with shading showing the 95% confidence interval. The fitted parameter k reflects each listener's overall sensitivity to monaural envelope correlation.

product of the spectral bandwidth of each band (W , in Hz) and the stimulus duration (D , in sec), which is proportional to the expected number of envelope modulation periods. Apart from the direct current, the average envelope power spectrum of a bandpass Gaussian noise falls linearly as a function of modulation rate, with a maximum rate corresponding to the width of the band (Dau *et al.*, 1999). The variable WD therefore represents the maximum possible number of modulation periods in a particular sample; the expected number of modulation periods is approximately $0.64WD$ (Rice, 1954). To examine whether the joint effects of stimulus bandwidth and duration are due to increased numbers of inherent modulation periods available to the listener, we fitted (via maximum-likelihood algorithm) the PC scores using the following multiple-look model:

$$PC = 100 * \phi\left(\frac{d'}{\sqrt{2}}\right), \quad (1)$$

where ϕ denotes the cumulative-normal function, and d' is the sensitivity index. According to the multiple-look model (Green and Swets, 1966; Viemeister and Wakefield, 1991), d' is related to WD by

$$d' = k\sqrt{WD}, \quad (2)$$

where k is the fitting parameter. The right column of panels in Fig. 1 shows listener data, re-plotted as a function of $0.64WD$, the expected number of envelope modulation periods per stimulus, and the data fit (solid lines, red online). The shaded area around each fit indicates the 95% confidence interval, derived from cumulative binomial probability functions, and values of k appear in the upper right of each panel. These fits capture some of the trends in the data but miss others. For example, for small numbers of modulation periods (<3.2), L1 performed better than predicted, whereas both L2 and L3 performed more poorly than predicted. In other words, the multiple-look model over-predicted the rate at which PC improves with increasing WD for L1, but under-predicted the rate of improvement for L2 and L3.

Listener responses in the extended data (not shown) indicate some reliable differences in listener responses across samples within a bandwidth/duration condition. The split half correlation, based on percent correct for each of the 25 correlated samples, was significant for all five listeners ($\alpha=0.05$), with values that ranged from $r=0.44$ (L5) to $r=0.95$ (L2). That is, within a listener, responses differed reliably across samples when both the correlated and random stimuli were frozen. Higher correlations could reflect either lower levels of internal noise or greater stability in listener strategy. There was also consistency between listeners. For example, L2 and L3 both provided extended data for the 100-Hz, 100-ms stimulus, and the pattern of percent correct by sample in these two sets of data was correlated at $r=0.79$ ($p < 0.001$). This high correlation suggests that the two listeners may have performed the task based on the same perceptual cue. The correlation associated the 100-Hz, 200-ms data provided by L3 and L4 was lower and non-significant ($r=0.29$, $p=0.080$).

The finding of consistent differences in listener responses to stimuli with a common bandwidth and duration indicates that WD does not fully capture the stimulus features underlying sensitivity to MECP, as WD is a single value for these stimuli. The failure of WD to account for listener responses on a stimulus-by-stimulus basis is not entirely surprising because the product of bandwidth and duration is proportional to the *average* number of modulation periods, but there is variability in the actual number of modulation periods across stimulus samples. An analysis was undertaken to see if variation in the actual number of modulation periods in each frozen sample of correlated noise within a bandwidth/duration condition was associated with listener responses. The equivalent modulation rate for each of the 100-Hz-wide low-frequency bands, including onset and offset ramps, was estimated by extracting the Hilbert envelope. The Hilbert envelope was then transformed into the envelope-frequency domain,

and the frequency associated with the midpoint in the cumulative power spectrum as a function of rate was identified. This estimate of the equivalent modulation rate had a mean of 59.9 Hz [standard deviation (s.d.)=9.3 Hz] for the 100-ms stimuli, and a mean of 65.3 Hz (s.d.=6.9 Hz) for the 200-ms stimuli. However, these estimates of equivalent modulation rate were only weakly correlated with the pattern of percent correct across samples, transformed to rationalized arcsine units (Studebaker, 1985). The correlation was significant for L2 ($r=0.44$, $p=0.014$, one tailed), but non-significant for all the other listeners, with correlations ranging from $r=0.31$ (L3) to $r=-0.03$ (L5). Because the inherent envelope fluctuation rate is linearly related to the number of modulation periods in each sample, the extended data provide only weak support for the hypothesis that MECF is limited by the number of modulation periods in the interval containing the correlated stimuli.

4. Discussion

The motivation for this experiment was to better understand why MECF becomes less accurate as stimulus bandwidth is reduced below 100 Hz. The hypothesis was that the limited number of envelope modulation periods available to the listener is responsible for poor performance at narrow bandwidths, such that increasing either bandwidth or duration by a constant factor has the same beneficial effect on performance. The data roughly conform to this expectation.

While increasing bandwidth and/or duration improved performance, there were extensive individual differences, even among this group of highly trained listeners. Data for the two poorest performers were highly variable. Of the remaining three listeners, data for two of them roughly conformed to the expected pattern of results, based on the square root of the number of envelope periods; one caveat was poorer than expected performance for stimuli with small numbers of envelope modulations. The best performing listener showed very modest effects of changing either duration or bandwidth, and was most sensitive to monaural envelope correlation overall. Compared to the model fit and to performance of other listeners, this listener performed particularly well for stimuli with small numbers of envelope modulation periods, even when stimuli contained fewer than one full modulation period.

Subjectively, L1 reported using different cues in different stimulus conditions. For most conditions, where the duration was relatively long, he reported that the cue associated with the correlated stimulus was roughness or a grinding quality. This cue was reportedly also used by the other four listeners. At short durations, L1 reported relying exclusively on a pitch cue. Such a cue might be related to differences in relative levels for two bands of random noise. When the envelope modulation rate is slow relative to the stimulus duration, a random-band pair could be dominated by energy in the low band or by energy in the high band. Under these conditions, one band might dominate the pitch of the pair, or two pitches could be heard. In contrast, pairs of correlated bands are equal in level, so the pitch is likely to be relatively constant from sample to sample. In the data of L1, the use of this pitch cue may be responsible for the relatively modest effects of changing either the bandwidth or the duration.

Listener responses to particular noise samples within a bandwidth/duration condition were relatively consistent across trials, accounting for as much as 90% of the variance in listener response (L2). Responses were unrelated or weakly related to the equivalent modulation rate (and therefore number of envelope fluctuation periods) of each sample, accounting for only up to 19% of the variance in listener response (L2). That is, for a particular bandwidth and duration, the number of modulation periods in the correlated stimulus was not a strong predictor of performance, despite the fact that WD was a fairly good predictor of performance across bandwidth and duration. It is unclear how to interpret this result, but one possibility is that the method used to estimate the mean equivalent modulation rate for each sample does not accurately reflect the perceptually dominant envelope rate or salient envelope features. Another possibility is that features of the random stimuli could have played a dominant role in

responses for the extended listening conditions, where correlated and random samples were frozen within a trial. While additional analyses could shed light on the best way to understand listener responses based on features of the frozen noise stimuli, enthusiasm for this approach was substantially reduced by the observation of marked individual differences between listeners.

Finally, although the general trends of the present dataset are consistent with a multiple-look model, wherein the number of envelope modulation periods predicts performance, it appears that this model does not capture all aspects of the data. While the multiple-look model over-predicted performance of L2 and L3 for small numbers of envelope periods, it under-predicted performance of L1 in these conditions. Subjective reports corroborate the idea that listeners are using different cues.

Acknowledgments

This work was supported by the National Institutes of Health, Grant No. NIDCD: R01 DC000418 (J.W.H.).

References and links

- ANSI (2010). *ANSI S3.6-2010, American National Standard Specification for Audiometers* (American National Standards Institute, New York).
- Buss, E., Hall, J. W., and Grose, J. H. (2013). "Monaural envelope correlation perception for bands narrower or wider than a critical band," *J. Acoust. Soc. Am.* **133**, 405–416.
- Dau, T., Verhey, J., and Kohlrausch, A. (1999). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," *J. Acoust. Soc. Am.* **106**, 2752–2760.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).
- Moore, B. C. J., and Emmerich, D. S. (1990). "Monaural envelope correlation perception, revisited: Effects of bandwidth, frequency separation, duration, and relative level of the noise bands," *J. Acoust. Soc. Am.* **87**, 2628–2633.
- Rice, S. O. (1954). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York), pp. 133–294.
- Richards, V. M. (1987). "Monaural envelope correlation perception," *J. Acoust. Soc. Am.* **82**, 1621–1630.
- Richards, V. M. (1988). "Components of monaural envelope correlation perception," *Hear. Res.* **35**, 47–57.
- Richards, V. M. (1989). "Comparing monotic, diotic, and dichotic presentation modes in synchrony detection," *J. Acoust. Soc. Am.* **85**, S143.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.