Target enhancement and noise cancellation in the identification of a rudimentary sound source in noise

Robert A. Lutfi^{a)} and Ching-Ju Liu

Auditory Behavioral Research Lab, University of Wisconsin, Madison, Wisconsin 53706
ralutfi@wisc.edu, chingjuliu@wisc.edu

Abstract: Perturbation analysis was used to determine the relative contribution of target enhancement and noise cancellation in the identification of rudimentary sound source in noise. In a two-interval, forced-choice procedure, listeners identified the impact sound produced by the larger of two stretched membranes as target. The noise on each presentation was the impact sound of a variable-sized plate. For four of five listeners, the relative weights on the noise were positive indicating enhancement, and for the remaining listeners, they were negative indicating cancellation. The results underscore the difficulty with evaluating models of masking solely in terms of measures of performance accuracy.

© 2011 Acoustical Society of America

PACS numbers: 43.66.B, 43.66.F [QJF]

Date Received: September 23, 2010 Date Accepted: December 07, 2010

1. Introduction

Computational models of human auditory detection and recognition in noise generally fall into three categories: those that implicate enhancement of the target, those that implicate cancellation of the noise, and those that implicate both (Durlach, 1963; Meddis and Hewitt, 1992; de Cheveigné, 1993; Piechowiak *et al.*, 2007). There have been few attempts to determine the relative contribution of these processes in studies, and no attempt, to the authors' knowledge, involving the identification of sound source attributes in noise. Coming closest to the later case is a literature on the identification of concurrent pairs of harmonic and inharmonic vowels. This work is reviewed by de Cheveigné *et al.* (1995) and would seem to support a predominant role of noise cancellation; however, the effects are small and are based on the measures of performance accuracy, which do not always best reflect the processes underlying identification. In the present study, we apply perturbation analysis to determine the relative contribution of target enhancement and noise cancellation in terms of the sign and magnitude of the listener's decision weights on the noise. The analysis is applied to a specific case in which the target is the impact sound corresponding to the larger of two stretched membranes, and the noise is the impact sound of a variable-sized plate. We begin by describing the specifics of the study so as to provide a concrete example for elucidating the logic of the approach.

2. Methods

2.1 Stimuli

The stimuli were approximations to the impact sounds of a stretched circular membrane (target) and a loosely suspended circular plate (noise). They were synthesized and presented via headphones using first-order analytic equations for the motion of these sources from standard acoustic texts (Morse and Ingard, 1968). The resulting impact sounds were a sum of exponentially damped sinusoids representing the individual partials of the sounds. For the membrane, the frequencies of the partials were

---

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

$f_{1M} \times [1.00 \ 1.594 \ 2.136 \ 2.296 \ 2.653 \ 2.918]$; for the plate, they were $f_{1P} \times [1.00 \ 2.80 \ 5.15 \ 5.98 \ 9.75 \ 14.09]$. The values of f_{1M} and f_{1P} varied from one presentation to the next, as will be described shortly. The power of the first partial was fixed across trials and equated for both sources; the decay modulus of the first partial was fixed at 0.2 s for the membrane and 0.4 s for the plate. Both the power and decay moduli of higher-numbered partials decreased from the first partial proportionally with frequency. So as to keep the trial sequence at a reasonable length, a 5-ms cosine-squared ramp was used to truncate the sounds after 1 s. Sounds were played at a 44,100-Hz sampling rate with 16-bit resolution using a MOTU 896 audio interface. From the interface, the sounds were buffered through a Rolls RA62c headphone amplifier and then delivered diotically to listeners over Beyerdynamic DT 990 headphones. A loudness balancing procedure was used (see Lutfi *et al.*, 2008) to calibrate the overall sound power to be approximately 70 dB sound pressure level (SPL) at the eardrum. Listeners were seated individually in a double-walled, independent atomic center (IAC) sound-attenuated chamber.

2.2 Procedure

The values of f_{1M} and f_{1P} in hertz varied independently and at random on each presentation, as would correspond to changes in the size of the membrane and plate. The specific values were

$$f_{1M} = 500 \cdot \exp(\sigma_M \cdot \text{jnd} \cdot z_M), \quad (1a)$$

$$f_{1P} = 250 \cdot \exp(\sigma_P \cdot \text{jnd} \cdot z_P), \quad (1b)$$

where $\text{jnd} = \log(1.002)$ is an estimate of the just-noticeable difference in frequency (Wier *et al.*, 1977) and z is a random normal deviate selected independently for the membrane and plate on each presentation. The value of σ_P was fixed at 10. In different conditions, σ_M took on values between 10 and 80, somewhat different for each listener. In the two-interval forced-choice procedure with feedback, listeners were instructed to select the interval containing the impact sound corresponding to the larger membrane (lower f_{1M}). Specific values of σ_M were selected individually for each listener so as to obtain a range of performance levels across all listeners from just above chance to near perfect performance. Conditions were run in random order on different days, each day consisting of a single 1-h experimental session of eight blocks of trials, 50 trials per block. A total of 400 trials were run for each condition. Listeners were two male and three female students of the University of Wisconsin—Madison, aged 24 to 36 yr. All had extensive previous experience with the impact sounds in similar two-interval, forced-choice tasks. The results of a standard hearing evaluation showed all to have normal hearing sensitivity of 15 dB hearing level (HL) or better from 250 Hz to 8 kHz (ANSI S3.6-1996).

2.3 Analysis

Consider first the predictions of noise cancellation models for this experiment. These models are comprised of two stages. The first stage involves a process of noise equalization or normalization across different observations of the noise; the second stage involves a process of subtraction of the observations which results in cancellation of the noise. In different versions of these models, the cancellation process is assumed to occur across noise observations made at different points in time (as in the old-plus-new heuristic of Bregman, 1990) across different ears (as in the binaural masking level difference model of Durlach, 1963) or across different frequency channels (as in the model of comodulation masking release model of Piechowiak *et al.*, 2007). In the present application, the cancellation must occur across time; specifically, across the two intervals of each trial. Let $T1$ and $T2$ represent the sizes of the target on the two intervals, and let $N1$ and $N2$ represent the corresponding sizes of the plate. The decision variable of the noise cancellation model is then of the form

$$D = (T1 + N1) - [(T2 + E) + (N2 + E)], \quad (2)$$

where E is the noise equalization term that is applied to both target and noise in the second interval of the trial, $E = a(N1 - N2)$ with $a \approx 1$. In practice, a fails to take on a value exactly equal to one due to assumed imperfections in the equalization process. Substituting the expression for E in Eq. (2) and rearranging terms yields

$$D = (T1 - aN1) - (T2 - aN2) = \Delta T - a\Delta N, \quad (3)$$

where Δ denotes the difference in target or noise across the two intervals. Equation (3) thus shows that noise cancellation requires a negative decision weight on the noise ($-a$) relative to that of the target. Target enhancement models, by comparison, attribute noise interference to imperfections in the enhancement process, which results in enhancement of certain elements of the noise as well as the target. The decision weight on the noise in these models is therefore positive and the decision variable has the form

$$D = (T1 + aN1) - (T2 + aN2) = \Delta T + a\Delta N, \quad (4)$$

with $0 < a \leq 1$. Filter models of masking are common examples of target enhancement models inasmuch as the elements of the noise that pass through the target filter are, like the target, given positive weight in the decision process (cf. [Patterson, 1976](#)).

To evaluate the models given by Eqs. (3) and (4), we perform a general linear regression on the trial-by-trial data in which the relation between the probability of a first interval response, $P(R = 1)$, and stimulus variables is given by

$$\text{logit}[P(R = 1)] = c_1\Delta T + c_2\Delta N + e, \quad (5)$$

where c_1 and c_2 are the estimated regression coefficients and e is the regression error assumed to result from various additive sources of internal noise (imperfections in the decision process). In practice, the values of c_1 and c_2 were obtained by regressing the listener's trial-by-trial response on the values of ΔT and ΔN using the glmfit routine of MATLAB v.7.0.1. These values were then used to compute a *relative* decision weight on the noise given by

$$w = \frac{c_2}{c_1 + |c_2|}, \quad (6)$$

where $c_1 > 0$. The empirical question is whether the obtained values of w are positive, indicating target enhancement, or negative, indicating noise cancellation. Note also, however, that Eq. (5) allows for the evaluation of a third alternative in which the noise simply serves to distract attention away from the target, without itself being given any weight (e.g., [Carlyon and Moore, 1986](#); [Werner and Bargones, 1991](#)). In this case, w is expected to be near zero with performance dictated primarily by internal noise, e , generated as a result of the distraction caused by the masker.

3. Results and conclusions

Across all conditions and listeners, the fits to the data provided by Eq. (2) were quite good. The deviance of the model was never more than twice the associated degrees of freedom, which by rule of thumb is taken to indicate that an alternative model with more free parameters would not provide a significantly better fit to the data ([Snijders and Bosker, 1999](#)). In what follows, then, we focus on the obtained values of w resulting from the regression. Figure 1 gives for each listener (different symbols) the obtained w and corresponding d' for the different values of σ_M (repeated symbols). Error bars give the 95% confidence intervals and dashed lines give the upper and lower bounds on the data corresponding to the best performance achievable given a particular value of w and zero internal noise (zero regression error). In interpreting these data, we note that points falling above the horizontal line ($w > 0$) are consistent with target enhancement, while points falling below this line ($w < 0$) indicate noise cancellation. The results give strong

Fig. 1. For different listeners (different symbols), the obtained values of w and d' are shown for the different values of σ_M (repeated symbols). Error bars give the 95% confidence intervals and dashed lines give the upper and lower bounds on the data corresponding to the best performance achievable given a particular value of w and zero internal noise. Points falling above the horizontal line ($w > 0$) indicate target enhancement, points falling below the horizontal line ($w < 0$) indicate noise cancellation, and points falling on the horizontal line indicate noise distraction wherein performance is dictated primarily by internal noise, e .

support to the target enhancement model, with the exception of one listener whose data are given by the yellow diamonds. Less than ideal performance of this listener reflects a comparison of target and noise for which a large T tends to appear smaller in the presence of a larger N , and visa versa. Note also that despite the clear difference in this listener's decision weights, the values of d' for this listener are distributed well within the range of values for the other listeners. Therefore, this listener represents a case where the analysis of performance accuracy alone would not have revealed the dramatic difference in decision strategy.

The fact that very different decision strategies can result in very similar identification performance in sound source identification tasks has recently been underscored by Lutfi and Liu (2007). These authors obtained decision weights from listeners to evaluate decision strategy in tasks involving the judgment of the material and size of rudimentary objects (bars, plates, and membranes), as well as the hardness of the striking mallet. Significant differences in decision strategy were found across listeners within each task, but identification accuracy was in most cases quite similar across listeners within tasks. The results of Lutfi and Liu (2007) and the dramatic example given in the present study suggest that care should be taken in evaluating the models of masking solely in terms of the predictions they make for measures of performance accuracy.

Acknowledgment

This research was supported by NIDCD Grant No. 5R01DC006875-05.

References and links

- ANSI S3.6-1996 (1996). *American National Standard Specification for Audiometers* (American National Standards Institute, New York).
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 213–393.

- Carlyon, R. P., and Moore, B. C. J. (1986). "Continuous versus gated pedestals and the 'severe departure' from Weber's law," J. Acoust. Soc. Am. **79**, 453–460.
- de Cheveigné, A. (1993). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.
- de Cheveigné, A., and McAdams, S., Laroche, J., and Rosenberg, M. (1995). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," J. Acoust. Soc. Am. **97**, 3736–3748.
- Durlach, (1963). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.
- Lutfi, R. A. and Liu, C. J. (2007). "Individual differences in source identification from synthesized impact sounds," J. Acoust. Soc. Am. **122**, 1017–1028.
- Lutfi, R. A., Liu, C. J., and Stoelinga, C. N. J. (2008). "Level dominance in sound source identification," J. Acoust. Soc. Am. **124**, 3784–3792.
- Meddis, R., and Hewitt, M. J. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies," J. Acoust. Soc. Am. **91**, 233–245.
- Morse, P. M., and Ingard, K. U. (1968). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ), pp. 175–191.
- Patterson, R. D. (1976). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
- Piechowiak, T., Ewert, S. D., and Dau, T. (2007). "Modeling comodulation masking release using an equalization-cancellation mechanism," J. Acoust. Soc. Am. **121**, 2111–2126.
- Snijders, T. A. B., and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Sage Publishers, London), p. 49.
- Werner, L. A., and Bargones, J. Y. (1991). "Sources of auditory masking in infants: Distraction effects," Percept. Psychophys. **50**, 405–412.
- Wier, C. C., Jesteadt, W., and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.