

Variable-duration notched-noise experiments in a broadband noise context

James J. Hant, Brian P. Strobe, and Abeer A. Alwan

*Speech Processing and Auditory Perception Laboratory, Department of Electrical Engineering,
School of Engineering and Applied Sciences, UCLA, 405 Hilgard Avenue, Los Angeles, California 90095*

(Received 27 April 1998; accepted for publication 30 June 1998)

A variable-duration notched-noise experiment was conducted in a noise context. Broadband noise preceded and followed a tone and notched noise of similar duration. Thresholds were measured at four durations (10, 30, 100, and 300 ms), two center frequencies (0.6, 2.0 kHz), and five relative notch widths (0.0, 0.1, 0.2, 0.4, 0.8). At 0.6 kHz, 10-ms thresholds decrease 6 dB across notch widths, while 300-ms thresholds decrease over 35 dB. These trends are similar but less pronounced at 2 kHz. In a second experiment, the short-duration notched noise was replaced with a flat noise which provided an equivalent amount of simultaneous masking and thresholds dropped by as much as 20 dB. A simple combination of simultaneous and nonsimultaneous masking is unable to predict these results. Instead, it appears that the elevated thresholds at short durations are dependent on the spectral shape of the simultaneous masker. © 1998 Acoustical Society of America.

[S0001-4966(98)02510-7]

PACS numbers: 43.71.An, 43.66.Ba, 43.66.Dc [RVS]

INTRODUCTION

Short-duration speech signals, such as plosives, are often confused in noisy environments. To model these confusions, it is necessary to characterize our ability to resolve the spectral components of short-duration stimuli in the context of background noise. Previous notched-noise experiments using a short-duration tone and long-duration notch have shown that frequency selectivity develops nearly instantaneously (Moore *et al.*, 1987; Wright and Dai, 1994). In this paper, a variable-duration notched-noise experiment is reproduced in a noise context. Specifically, the tone and notched noise are of similar duration and surrounded in time by long-duration (300-ms) white noise. Tone thresholds are therefore determined by the combination of three potential maskers: the preceding white noise, the simultaneous notched noise, and the trailing white noise.

Previous studies suggest that a combination of maskers may result in masking which is greater than that predicted by a linear sum of each individual masker. These studies have shown masking increases of 7 dB for combinations of simultaneous maskers (Green, 1967), 15 dB for combinations of nonsimultaneous maskers (Oxenham and Moore, 1995), and 14 dB for combinations of simultaneous and nonsimultaneous maskers (Jesteadt, 1996). This “excess masking” can be predicted by power-law models in which the effects of the individual maskers are compressed before they are added together (for a review see Humes and Jesteadt, 1989).

In this paper, the variable-duration notched-noise data are analyzed in terms of either a linear or nonlinear combination of simultaneous and nonsimultaneous masking. To further assess this approach, a second experiment is conducted in which the notched noise is replaced by a flat noise which provides an equivalent amount of simultaneous masking.

I. EXPERIMENT 1

A. Stimuli

Figure 1 shows a schematic spectrogram of the stimuli used in the experiment. The notched-noise masker and tone signal were presented between two 300-ms white-noise segments. White noise was generated by randomly choosing 1 of 32 digitally synthesized Gaussian noise sequences. The notched noise was created by filtering the white noise with an 801-tap bandstop filter. Both the notched noise and the surrounding white noise had a spectrum level of 36 dB/Hz and were generated from different Gaussian noise sequences.

To reduce the effect of spectral splatter, all stimuli were gated on and off using raised-cosine windows with rise/fall times of 10 ms. Signal durations of 10, 30, 100, and 300 ms were defined from the half-rms level (i.e., -6 dB). Figure 2 shows a diagram of the temporal relationships between the onsets and offsets for the tone, notch, and white noise. The notched noise was centered (in time) around the tone and was 20 ms longer than the tone. This ensured that the full amplitude of the notch was in place throughout the entirety of the tone. To maintain a constant spectrum level across time in the frequency region outside the notch, the transients of the notched-noise masker were opposite to those of the white-noise masker.

The experiment was conducted at tone frequencies of 0.6 and 2 kHz. All notches were symmetric (in frequency) around the tone and characterized by $\Delta f/c_f$, where c_f is the tone frequency, and Δf is the frequency difference between the tone and either edge of the notch. The skirts of the spectral notch had slopes which ranged from -25 dB/50 Hz (at a tone duration of 10 ms) to -60 dB/50 Hz (at a tone duration of 300 ms).

Thresholds were measured for $\Delta f/c_f$ equal to (0.0, 0.1, 0.2, 0.4, 0.8) which correspond to notch widths of (0.0, 0.12, 0.24, 0.48, 0.96) kHz at a center frequency of 0.6 kHz, and

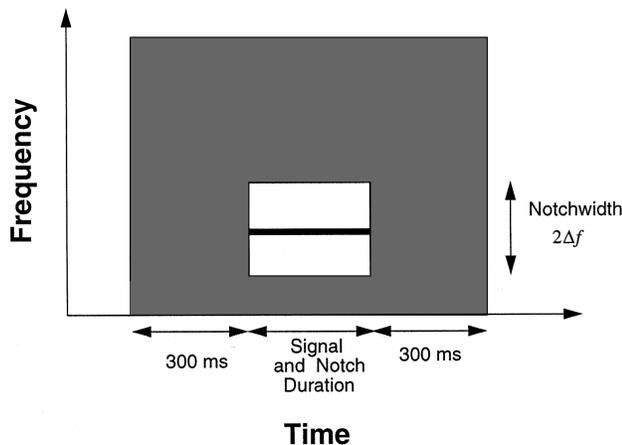


FIG. 1. Schematic spectrogram of the stimuli in experiment 1. The notched noise and tone (of a specific duration) are surrounded by two 300-ms white-noise maskers. The notch is symmetric around the tone and defined in terms of the fractional deviation from the tone's frequency ($\Delta f/c_f$).

(0.0, 0.4, 0.8, 1.6, 3.2) kHz at a center frequency of 2 kHz. To measure the amount of nonsimultaneous masking by the surrounding white noise, additional data points were collected using a wide notch width from 50 to 7950 Hz.

B. Subjects

Five audiometrically normal subjects (two male, three female) participated in the experiments. The first two authors participated as subjects and were experienced listeners in psychoacoustic experiments. The other three subjects were paid and had no previous experience as laboratory listeners.

C. Protocol

Stimuli were presented binaurally to listeners in a sound attenuating room via Telephonics TDH49P headphones. The ringing of the headphones was measured to be negligible at the frequencies and durations used in the experiment. Computer software generated the test tokens as 16 bit/16 kHz digital numbers. An Ariel ProPort 656 board performed digital-to-analog conversion. The resulting analog waveforms were amplified using the pre-amp of a Sony 59ES DAT recorder, which was connected to the headphones. The entire system was calibrated within ± 0.5 dB before each experiment using a Larson Davis 800B sound level meter.

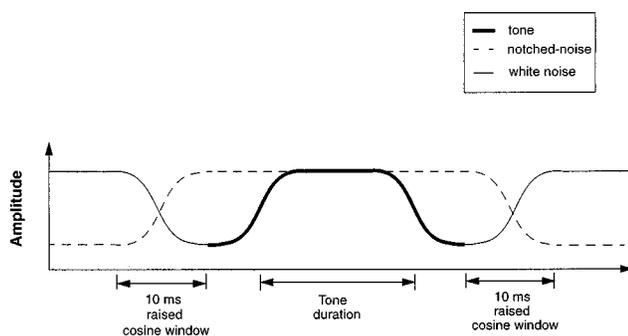


FIG. 2. Schematic of the temporal transitions for the tone, notched noise, and white noise. All stimuli are turned on and off with a raised cosine window of length 10 ms. Durations are defined from the half-rms levels.

Masked thresholds were determined using an adaptive 2I, 2AFC paradigm with no feedback (Levitt, 1971). Three correct responses determined a successful subtrial while one incorrect response determined an incorrect subtrial. Thresholds, therefore, are defined to be the 79% correct point. Step sizes were initially set to 4 dB, then reduced to 2 dB after the first reversal, and finally to 1 dB after the third reversal. From a total of nine reversals, the average of the last six determined threshold for each trial. The mean of two trials determined the final threshold. If the difference between the two trials was greater than 3 dB, a third point was taken and the median of all three trials determined the final threshold. Thresholds for the wide notch width condition were determined from one trial.

Subjects were trained for one hour before beginning the experiments. No training effects were apparent in the final data. One subject had particular difficulty detecting tones at the narrow notch widths across all durations. Despite substantial training, thresholds *increased* by nearly 5 dB between the 0.0 and 0.1 notch width conditions. Data from this subject were not included in the analyses below.

II. RESULTS AND DISCUSSION

To model the results of experiment 1, the masking contributions of the preceding white noise, the simultaneous notched noise, and the trailing white noise are considered. The results of notched-noise experiments using a short-duration tone and long-duration notch have shown that frequency selectivity develops nearly instantaneously (Moore *et al.*, 1987; Wright and Dai, 1994). The difference between these studies and experiment 1 is the noise context which adds forward and backward masking noise. Therefore, in the following analyses, the amount of simultaneous masking is determined by a traditional filtering model which uses auditory filters derived from long-duration notched-noise experiments (Glasberg and Moore, 1990). The masking contributions of the surrounding white noises are lumped together into a single nonsimultaneous-masking term. Tone thresholds are then predicted as either a linear or nonlinear combination of simultaneous and nonsimultaneous masking.

A. Predictions of a linear model

Assuming a linear combination of simultaneous and non-simultaneous masking, thresholds from experiment 1 are predicted by the following equation:

$$T_{\text{tot}} = 2K \int_b^{\infty} N(g)W(g)dg + T_{\text{ns}}, \quad (1)$$

where T_{tot} is the threshold prediction, g is the normalized frequency deviation from the tone center frequency, K is the threshold SNR due to simultaneous masking, b is the relative notch width ($\Delta f/c_f$), $N(g)$ is the power density of the noise, $W(g)$ is the shape of the auditory filter centered around the tone, and T_{ns} is the tone threshold due to nonsimultaneous masking.

Auditory filters are assumed to have the shape of a symmetric roex function,

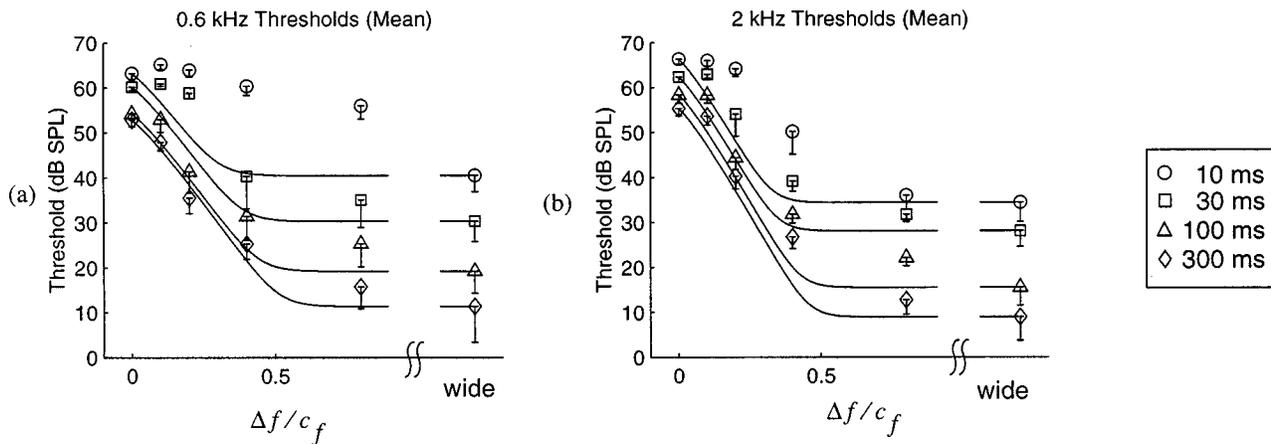


FIG. 3. Mean notched-noise thresholds. (a) Thresholds of 0.6-kHz tones as a function of the relative notch width, $(\Delta f/c_f)$, with signal duration as a parameter. Error bars, shown below the data points, represent standard deviations across subjects. Wide notch width thresholds, corresponding to a notch between 50 and 7950 Hz, are denoted by “wide” on the horizontal axis. (b) Thresholds of 2-kHz tones. The lines represent predictions of the linear model at each duration.

$$W(g) = (1-r)(1+pg)e^{-pg} + r, \quad (2)$$

where p specifies the slope of the filter skirts and therefore sets the filter bandwidth, and r determines the dynamic range of the filter. The filter bandwidths used were from Glasberg and Moore (1990), and r was set to -75 dB. Here, T_{ns} and K are determined from the wide and 0.0 notch thresholds, respectively.

Figure 3(a) plots mean thresholds of the 0.6-kHz tone as a function of the relative notch width $(\Delta f/c_f)$ with signal duration as a parameter. Standard deviations are expressed by error bars and, for clarity, are only shown below the average thresholds. The mean data at 2 kHz are shown in Fig. 3(b). Wide notch-width thresholds, corresponding to a notch from 50 to 7950 Hz, are denoted by “wide” on the horizontal axes of Fig. 3(a) and (b). The solid lines represent the predictions of the linear model for each duration.

At short durations (10 and 30 ms), a linear combination of simultaneous and nonsimultaneous masking underestimates measured thresholds. The discrepancy is most notable for the 10-ms thresholds at 0.6 kHz, but the trend exists throughout the data. Between the 0.0- and wide-notch conditions, the combination of masking is greater than the (linear) sum of its parts. At the longer durations (100 and 300 ms), the contribution of the nonsimultaneous masker is reduced and the linear model provides better threshold predictions.

B. Predictions of a nonlinear model

Auditory compression has been used in nonlinear models that predict combinations of forward and backward masking (Penner, 1980; Oxenham and Moore, 1994, 1995) and combinations of simultaneous and nonsimultaneous masking (Jesteadt *et al.*, 1996). Nonlinear auditory models which include adaptation (Zwislocki, 1969; Dau *et al.*, 1996) further imply time-varying compression. To predict the current data, we assume summation of simultaneous and nonsimultaneous masking after a power-law compression which, for generality, is allowed to vary as a function of the stimulus duration. The total threshold, T_{tot} , is the sum of the thresholds due to

simultaneous masking, T_s , and nonsimultaneous masking, T_{ns} , after each term has been raised to a power, q :

$$(T_{tot})^q = (T_s)^q + (T_{ns})^q, \quad (3)$$

where

$$T_s = 2K \int_b^\infty N(g)W(g)dg. \quad (4)$$

The parameters T_{ns} , q , and the scaling factor K are iteratively optimized to reduce the mean squared error (in dB) of the model predictions for each duration. As with the linear predictions above, the logarithm of T_{tot} is used to provide threshold predictions on a dB scale.

Figure 4(a) and (b) shows the predictions of the nonlinear model at tone frequencies of 0.6 and 2 kHz, respectively. While the predictions are better than those from the linear model, the error is still considerable, especially at short durations. For both center frequencies, the optimal q values decrease with shorter durations, implying an increase in the nonlinear summation of masking.

III. EXPERIMENT 2

A simple modification of the first experiment was used to assess the utility of the current modeling approach. The simultaneous notched noise was replaced with a flat-noise masker (0–8 kHz). Figure 5 shows a schematic spectrogram of the stimuli. Assuming traditional measures of frequency selectivity, the level of the simultaneous flat noise was set to provide the same amount of simultaneous masking as the notched noise in experiment 1. With identical amounts of simultaneous and nonsimultaneous masking, we expect similar thresholds.

For a given filter shape $W(f)$ (Glasberg and Moore, 1990), and relative notch width $(\Delta f/c_f)$ equal to b , the spectrum level of the noise, N_{eq} , which equates the amount of simultaneous masking in experiments 1 and 2, is given by the following equation:

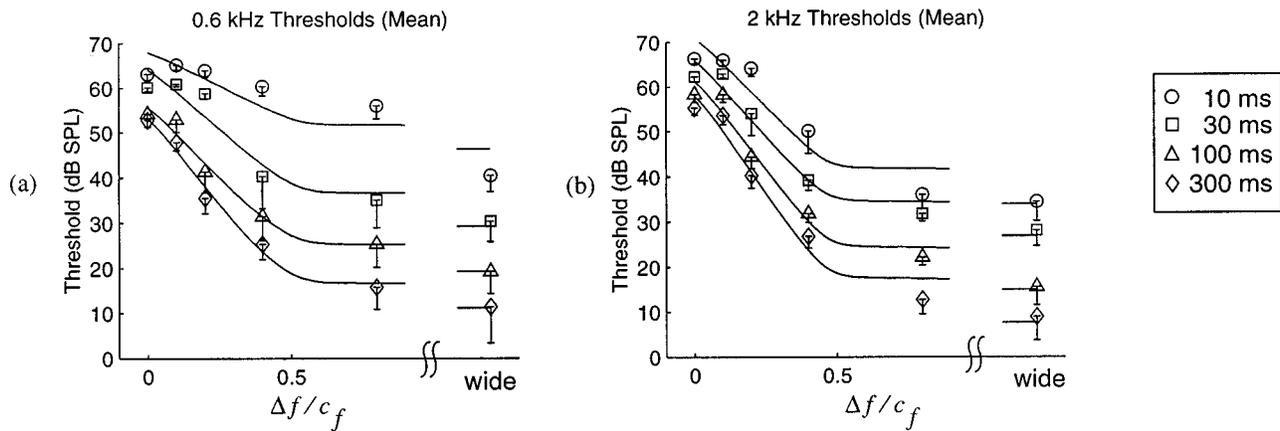


FIG. 4. Mean notched-noise thresholds. Same as Fig. 3 except the lines represent predictions of the nonlinear model.

$$N_{eq} = \frac{N_0 [\int_0^{(1-b)c_f} W(f) df + \int_{(1+b)c_f}^8 \text{kHz} W(f) df]}{\int_0^8 \text{kHz} W(f) df}, \quad (5)$$

where N_0 is the spectrum level of the notched noise (36 dB/Hz).

The four subjects whose data were analyzed in experiment 1, participated in experiment 2. Tone thresholds were measured for equivalent noise levels, N_{eq} , corresponding to notches of 0.2, 0.4, and 0.8. The onsets and offsets of the surrounding noise, tone, and simultaneous noise masker were identical to those used in experiment 1. Tone thresholds were measured using the same procedure as in the first experiment. Measurements were made only at a 10-ms tone duration.

IV. COMPARING RESULTS

In Fig. 6, thresholds from experiment 2 (squares) are compared to the corresponding notched-noise thresholds from experiment 1 (circles) and the nonlinear threshold predictions (solid lines). For both experiments, thresholds are averaged across the four subjects with error bars representing standard deviations. The values of N_{eq} (in dB/Hz) are shown on the horizontal axes below the corresponding notch.

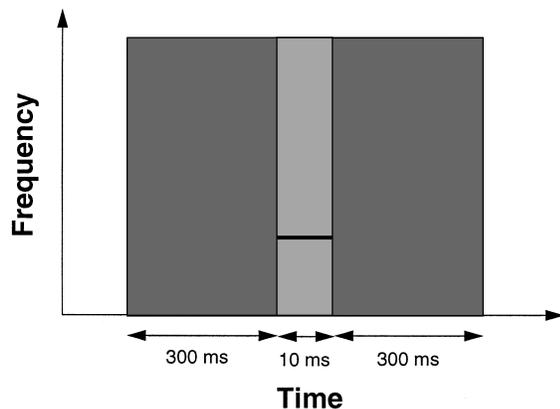


FIG. 5. Schematic spectrogram of the stimuli in experiment 2. A 10-ms tone and a flat, simultaneous masker of spectrum level, N_{eq} , are surrounded by two 300-ms white-noise maskers, each with a spectrum level of 36 dB/Hz. Tone frequencies are either 0.6 kHz or 2 kHz.

Thresholds with a flat-noise simultaneous masker are lower than those with a notched-noise masker. Despite a similar combination of simultaneous and nonsimultaneous masking, thresholds for experiment 2 are as much as 20 dB less than those for experiment 1. Therefore, no simple combination of simultaneous and nonsimultaneous masking predicts both results.

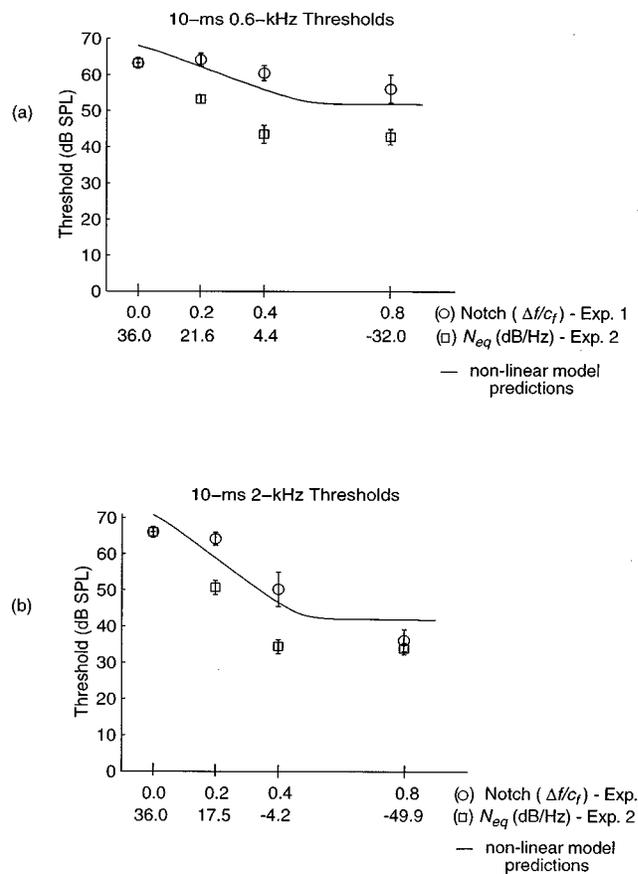


FIG. 6. Mean results of experiments 1 and 2: (a) 0.6-kHz, 10-ms tone thresholds for experiments 1 and 2 are denoted by the circles and squares, respectively. Thresholds are averaged across four subjects with error bars representing standard deviations. The values of N_{eq} (in dB/Hz) are shown on the horizontal axis below the corresponding notch. The solid line shows the threshold predictions of the nonlinear model. (b) Same as (a), at a center frequency of 2 kHz.

V. DISCUSSION

Our goal is to model the perception of short-duration speech sounds in a noise context. Toward that end, a traditional measure of frequency selectivity, the notched-noise experiment (Patterson, 1976), was reproduced in a noise context and at varying durations. The forward- and backward-masking components introduced by the surrounding noise were found to increase thresholds, especially at shorter durations, wider notch widths, and the lower center frequency. In an attempt to evaluate the suitability of “standard” psychoacoustic models, threshold predictions from a traditional filtering model with either a linear or nonlinear combination of simultaneous and nonsimultaneous masking were fit to these data. In a second experiment, each notched noise was replaced by a flat noise that was assumed to provide the same simultaneous masking. Thresholds were much lower with the flat noise, implying that no simple combination of masking can predict results from both experiments.

Instead, we are left to conclude that the spectral shape of the simultaneous masker influences the interaction with nonsimultaneous maskers. This result is not inconsistent with previous work that parameterized the perception of short-duration sounds in broadband noise as a reduction of usable frequency selectivity (Hant *et al.*, 1997). With such a reduction, threshold predictions for experiment 1 increase, as the tone is not as well separated from the simultaneous notch. Similarly, if “overly sharp” frequency selectivity estimations were used to set the noise level in experiment 2, then the amount of simultaneous masking would be lower than in experiment 1, and lower thresholds would be expected. Another pilot experiment using a narrow-band noise signal and a two-tone simultaneous masker, also surrounded in a noise context, provided a similar result.

One difficulty is defining a mechanism that reproduces these trends. An interaction between suppression and nonsimultaneous masking might appear likely. However, an interaction between the spectral shape of the simultaneous masker and a backward masker has been measured (see Fig. 12 in Hant *et al.*, 1997). Figure 7 replots these data with those from experiment 1 and shows similar trends. If suppression is considered nearly instantaneous, then there is little time for interaction with an extremely weak backward masker.

These results might still be partially attributed to an interaction of peripheral mechanisms, if 2-D differences in time-varying excitation patterns produced by an auditory model with adapting compression are considered. Similarly, at a higher level in the auditory system, thresholds might be predicted if we consider masking as the corruption of an evolving statistical estimate of the signal. Shorter durations and surrounding noise may hinder such statistical estimates.

Another strong possibility is that auditory perception is more readily characterized by the response to *changes* across time and frequency, than by the response to stationary stimuli. If so, then the extent of coherent changes across one dimension could influence the ability to detect changes in the other. For example, in experiment 1, as the duration of the notch increases, the ability to detect the spectral change

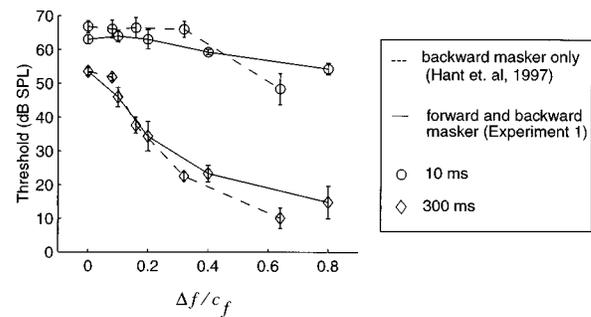


FIG. 7. Replot of the results of experiment 1 with those of a pilot notched-noise experiment from Hant *et al.* (1997). The pilot experiment was similar to experiment 1 except that the tone and notch were only followed by backward-masking noise. Thresholds from the pilot experiment, which was conducted at a center frequency of 1 kHz and relative notch widths of 0.08, 0.16, 0.32, and 0.64, are denoted by the dashed lines. 0.6-kHz thresholds from experiment 1 are denoted by the solid lines. Tone (and notch) durations are delineated by the circles (10 ms) and diamonds (300 ms). Thresholds for both experiments, are averaged across the same two subjects (JH and BS), with standard deviations represented by the error bars.

caused by the presence of the tone improves, and thresholds drop.

A similar approach may explain the discrepancy between experiments 1 and 2. In experiment 2, transients preceding and following the tone extended over a wider frequency range than in experiment 1. Well-correlated wide-bandwidth transients may have cued subjects to the low-energy valley of the masker, helping them locate “when” to listen for the tone and thus, lowering thresholds (Buus, 1985). However, in the 0.6-kHz 0.8-notch condition, large differences in the thresholds of experiments 1 and 2 are measured. Previous CMR data (Hall *et al.*, 1984) do not support significant decreases in threshold when the transient bandwidth increases beyond that of the 0.8 notch. Regardless, interpreting the current data in terms of “listening to the valleys” is difficult because the “valleys” in experiments 1 and 2 have different depths and unlike previous CMR studies, there’s only one “valley.” More experiments are necessary.

Predicting the discrepancy between our results and those of Moore *et al.* (1987) and Wright and Dai (1994) also remains a significant challenge. Clearly, the addition of nonsimultaneous noise (in the spectral region of the signal) degrades our ability to perceive short-duration signals. However, the nature and time course of this degradation are still not known. Additional psychoacoustic measurements, both in a noise context and using a wider range of time-varying stimuli, will help characterize what appears to be a time-varying process and should constrain the modeling possibilities.

We can conclude, however, that a classical filtering model with either a linear or nonlinear combination of simultaneous and nonsimultaneous masking does not directly predict the current data. These models are therefore inadequate to predict the perception of short-duration speech signals in a noise context.

ACKNOWLEDGMENTS

We thank our subjects for their cooperation. We would also like to thank R. Shannon and two anonymous reviewers

for their helpful comments. This work was supported in part by NIH-NIDCD Grant No. 1 R29 DC 02033-01A1 and by the Whitaker Foundation.

- Buus, S. (1985). "Release from masking caused by envelope fluctuations," *J. Acoust. Soc. Am.* **78**, 1958–1965.
- Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched noise data," *Hearing Res.* **47**, 103–138.
- Green, D. M. (1967). "Additivity of masking," *J. Acoust. Soc. Am.* **41**, 1517–1525.
- Dau, T., Puschel, D., and Kohlrausch, A. (1996). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Am.* **99**, 3615–3622.
- Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **76**, 50–56.
- Hant, J., Strobe, B., and Alwan, A. (1997). "A psychoacoustic model for the noise masking of voiceless plosive bursts," *J. Acoust. Soc. Am.* **101**, 2789–2802.
- Humes, L., and Jesteadt, W. (1989). "Models of the additivity of masking," *J. Acoust. Soc. Am.* **85**, 1285–1294.
- Jesteadt, W., Neff, D. L., and Kessler, C. J. (1996). "Combinations of forward and backward masking revisited," *J. Acoust. Soc. Am.* **100**, 2593.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Moore, B. C., Poon, P. W., Bacon, S., and Glasberg, B. (1987). "The temporal course of masking and the auditory filter shape," *J. Acoust. Soc. Am.* **81**, 1873–1880.
- Oxenham, A. J., and Moore, B. C. (1994). "Modeling the additivity of nonsimultaneous masking," *Hearing Res.* **80**, 105–118.
- Oxenham, A. J., and Moore, B. C. (1995). "Additivity of masking in normally and hearing impaired subjects," *J. Acoust. Soc. Am.* **96**, 1921–1934.
- Patterson, R. D. (1976). "Auditory filter shapes derived by noise stimuli," *J. Acoust. Soc. Am.* **59**, 640–654.
- Penner, M. J. (1980). "The coding of intensity and the interaction of forward and backward masking," *J. Acoust. Soc. Am.* **67**, 608–616.
- Wright, B., and Dai, H. (1994). "Detection of unexpected tones with short and long durations," *J. Acoust. Soc. Am.* **95**, 931–938.
- Zwislocki, J. J. (1969). "Temporal summation of loudness: An analysis," *J. Acoust. Soc. Am.* **46**, 431–441.