Durations required to distinguish noise and tone: Effects of noise bandwidth and frequency[a)]

Armin Taghipour[b)]

International Audio Laboratories Erlangen, a joint institution of Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits (IIS), Am Wolfsmantel 33, 91058, Erlangen, Germany

Brian C. J. Moore

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Bernd Edler

International Audio Laboratories Erlangen, a joint institution of Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits (IIS), Am Wolfsmantel 33, 91058, Erlangen, Germany

Perceptual audio coders exploit the masking properties of the human auditory system to reduce the bit rate in audio recording and transmission systems; it is intended that the quantization noise is just masked by the audio signal. The effectiveness of the audio signal as a masker depends on whether it is tone-like or noise-like. The determination of this, both physically and perceptually, depends on the duration of the stimuli. To gather information that might improve the efficiency of perceptual coders, the duration required to distinguish between a narrowband noise and a tone was measured as a function of center frequency and noise bandwidth. In experiment 1, duration thresholds were measured for isolated noise and tone bursts. In experiment 2, duration thresholds were measured for tone and noise segments embedded within longer tone pulses. In both experiments, center frequencies were 345, 754, 1456, and 2658 Hz and bandwidths were 0.25, 0.5, and 1 times the equivalent rectangular bandwidth of the auditory filter at each center frequency. The duration thresholds decreased with increasing bandwidth and with increasing center frequency up to 1456 Hz. It is argued that the duration thresholds depended mainly on the detection of amplitude fluctuations in the noise bursts.

I. INTRODUCTION

Perceptual audio coders (Bosi *et al.*, 1997; Brandenburg and Bosi, 1997; Brandenburg and Stoll, 1994; Stoll and Brandenburg, 1992) exploit the masking properties of the human auditory system to reduce the bit rate of digital recording and transmission systems; the audio signal is treated as the masker and quantization noise as the "probe" that is to be masked at the output of the decoder. In perceptual audio coders, the audio signal is split into brief segments, called "frames." The frame length may be fixed or it can vary according to the characteristics of the signal. Within each frame, the signal is filtered into a number of adjacent frequency bands called subbands. The signal in each subband is represented as a sequence of binary digits (bits). The greater the number of bits, the lower is the quantization noise relative to the audio signal. The most efficient coding, i.e., the minimum number of bits required to represent the audio signal in the subband without audible artifacts, is achieved if the quantization noise lies just below the masked threshold. A psychoacoustic model is used to estimate the masked threshold of the quantization noise and to allocate the number of bits to be used for that subband for that frame. Generally, the quantization noise in a given subband covers the whole spectral range of that subband, but the audio signal often has a narrower bandwidth (Bosi and Goldberg, 2003). The effectiveness of the audio signal in masking the quantization noise depends on the bandwidth of the audio signal and especially on whether it is tone-like or noise-like. This property is referred to as "tonality." Generally, a tone-like masker is less effective than a noise-like masker (Gockel *et al.*, 2002; Hall, 1997; Hellman, 1972; Verhey, 2002).

Both perceptually and physically, it takes some time to make a decision about the tonality of a sound; a very short burst of narrowband noise and a tone burst of the same duration and with the same center frequency sound very similar and are physically very similar. The main physical

---

difference is that the envelope of a narrowband noise fluctuates over time, but for short-duration signals, this fluctuation can be hard to detect (Stone *et al.*, 2008). Hence to exploit the fact that quantization noise is masked more effectively by a noise-like audio signal than by a tone-like audio signal, the frame length used in a perceptual coder must be sufficiently long. However, the required duration may vary depending on the center frequency and on the width of the subbands. As the width of a subband is increased, the amplitude fluctuations in a noise-like audio signal become more rapid (Bos and de Boer, 1966), and this would be expected to decrease the duration required to distinguish a tone from a noise. Although data are available on the detection of sinusoidal amplitude modulation as a function of stimulus duration (Sheft and Yost, 1990), we are not aware of any previous study that has measured the duration required to discriminate a tone from a narrowband noise of the same center frequency. Hereafter, this duration is referred to as a "duration threshold." The main goal of the present experiments was to estimate duration thresholds as a function of center frequency and of the bandwidth of the noise. The data were intended to be useful in improving the design of perceptual coders by indicating whether the frame length should vary with subband center frequency and bandwidth.

In experiment 1, duration thresholds were measured for isolated noise and tone bursts (Taghipour *et al.*, 2013). This provided baseline data for a relatively simple situation. However, in a perceptual coder, the decision about tonality has to be made for each frame, and the decision for a given frame can be influenced by the stimuli in preceding and following frames. For example, it is easier to detect a brief irregularity embedded in a regular or steady sound than it is to detect a brief regularity embedded in an irregular sound (Chait *et al.*, 2007; Pollack, 1968). In experiment 2, duration thresholds were measured for tone and noise segments embedded within longer tone pulses (Taghipour *et al.*, 2014). This corresponds to a situation where performance is expected to be relatively good because the presence of an embedded noise burst is indicated by a transition from regularity to irregularity. The use of this embedding reduced spectral broadening effects associated with short-duration stimuli, although spectral "splatter" could potentially provide a cue for detection of the noise burst (Taghipour *et al.*, 2013); the possible influence of spectral splatter is discussed later. The methods used in the two experiments differed somewhat because the experiments were conducted independently.

II. EXPERIMENT 1

In experiment 1, the duration required for subjects to distinguish between tone and noise bursts was estimated as a function of center frequency and of the bandwidth of the noise (Taghipour *et al.*, 2013).

A. Method

1. Stimuli and equipment

In each trial, two stimuli were presented consecutively with a silent gap of 800 ms between them. One of the following pairs was selected randomly for each trial: tone-tone, tone-noise, noise-tone, or noise-noise. The stimuli were gated with raised-cosine ramps. Because the duration of the stimuli was the independent variable in a run, the lengths of the ramps also varied. For overall durations (including ramps) up to 5 ms, the duration of each ramp was 1 ms; for overall durations between 5 and 10 ms, the duration of each ramp was 2 ms; beyond that, 3 ms ramps were used. For very short durations, the spectra of the stimuli broadened, and this might have led to audible spectral "splatter" (Taghipour *et al.*, 2013). This is discussed in more detail later. The estimated level of the stimuli at the eardrum was 75 dB sound pressure level (SPL) (see following text for details of calibration). This level was chosen on the basis of pilot experiments as it led to a comfortable loudness, given the short duration of the stimuli at the threshold for discrimination (duration thresholds were typically below 18 ms).

The noise bandwidth was specified on the ERB_N-number scale (Glasberg and Moore, 1990), which has units Cams (Moore, 2012). This scale was used because of its direct link to the bandwidth of the auditory filters. The bandwidth was 0.25, 0.5, or 1 Cams. The center frequencies were 345, 754, 1456, and 2658 Hz, corresponding to approximately 8.5, 13.5, 18.5, and 23.5 Cams. Because all stimuli had a bandwidth of 1 Cam or less, the characteristics of the stimuli at the output of an auditory filter at the same center frequency as the stimuli would have been similar to the characteristics of the stimuli themselves. The tone stimuli were generated deterministically, and their waveforms started at a positive-going zero-crossing. Different random noise stimuli were generated for each trial. For this purpose, white Gaussian noise was digitally filtered in the time domain by fourth order bandpass Butterworth filters. The total energy of each noise burst was adjusted to equal that of the tone burst.

The listening room had a background noise level of 25 dBA. Calibration of levels at the eardrum was done by means of an artificial head (KEMAR, GRAS, Holte, Denmark). Stimuli were computed digitally with a sample frequency of 48 kHz and a resolution of 16 bits. An RME Babyface digital-to-analog converter (Haimhausen, Germany) was used for playback. The stimuli were presented diotically via a pair of open electrostatic Stax SR-507 headphones with a SRM-600 driver unit (Saitama prefecture, Japan).

2. Procedure

Subjects were asked whether the two signals were the same or different. They were told that the signals would be the same on half of the trials and different on the other half. A response was counted as correct when the subject responded "same" and the two signals were the same (either tone-tone or noise-noise) or they responded "different" and the signals were different (either tone-noise or noise-tone). Otherwise, the response was counted as incorrect. Feedback in the form of a green or red light was provided after each trial via a graphical user interface indicating a correct or an incorrect response, respectively. A 3-down/1-up adaptive method was used that tracks the threshold corresponding to 79% correct (Levitt, 1971). Based on the outcome of a prior

J. Acoust. Soc. Am. **139** (5), May 2016

Taghipour *et al.* 2483

study (Taghipour *et al.*, 2012) and a pilot experiment, a fixed step size of 1 ms was chosen. Two randomly chosen conditions were presented interleaved, as suggested by Levitt (1971). The starting duration was between 13 and 21 ms. A run continued until 11 reversals for each condition had occurred. The average duration at the last six reversal points for a given condition was taken as the threshold estimate for that condition. A single threshold estimate was obtained for each condition and subject.

3. Subjects

Thirty self-reported normal-hearing subjects were tested. The "modified Thompson Tau test" and "Dixon's Q test" were used to check for outliers. Both tests revealed outliers for one or more conditions for three subjects, and all data for these three subjects were excluded from further analyses. Thus the final statistical analysis was based on thresholds for 27 subjects: 16 males and 11 females. They were aged between 20 and 49 yr (mean 28 yr, median 27 yr).

4. Design

Each subject was tested in three sessions of about 20–25 min each, separated by at least half a day. Prior to the main experiment, subjects read a page of instructions. They had 2–3 min of training by listening to signal pairs for which they were informed as to which pairs were identical and which were different. A training session of two runs with center frequencies 345 and 1456 Hz and bandwidths of 0.5 and 1 Cam, respectively, followed. This preparation/training phase took 15–20 min. In the main experiment, the conditions were presented in a random order.

B. Results

The variability of the thresholds was approximately proportional to the threshold values, so geometric mean thresholds across subjects were calculated. The mean duration thresholds across the 27 subjects and their 95% confidence intervals (assuming normally distributed data) are illustrated in Fig. 1. Duration thresholds are plotted on a logarithmic scale as a function of center frequency (also on a logarithmic scale) with bandwidth as parameter. Shapiro–Wilk tests of normality showed that the logarithms of the thresholds were normally distributed for six conditions but deviated somewhat from normality for the other six conditions. Because analysis of variance (ANOVA) is robust to moderate deviations from normality, a two-way repeated-measures ANOVA was conducted on the logarithm of the duration thresholds with factors bandwidth and frequency. There were significant effects of noise bandwidth [$F_{(2, 52)} = 106.4$, $p < 0.001$] and frequency [$F_{(3, 78)} = 56.2$, $p < 0.001$]. There was no significant interaction; [$F_{(6, 156)} = 1.0, p > 0.1$].

Post hoc pairwise comparisons (here and elsewhere based on Fisher's protected least significant difference, LSD, test) were conducted to investigate the effects of bandwidth and center frequency. Duration threshold decreased with increasing bandwidth (all pairs $p < 0.001$) and with
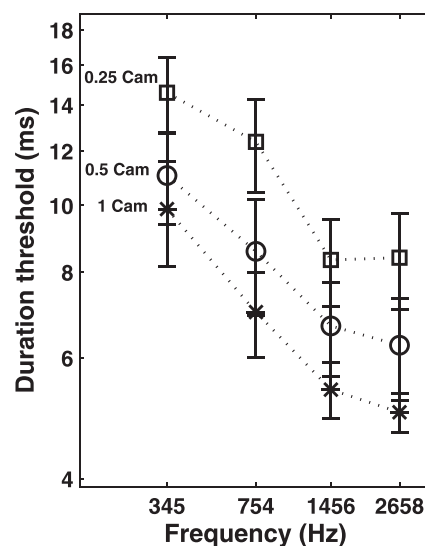

FIG. 1. Results of experiment 1: means and 95% confidence intervals of the duration thresholds across the 27 subjects are plotted as a function of frequency with noise bandwidth in Cams as parameter. Duration thresholds are shown in ms.

increasing frequency up to 1456 Hz (all pairs $p < 0.001$). No significant difference was found between the two highest center frequencies ($p = 0.405$).

When bandwidths were expressed in Cams, as in the preceding text, the results showed significant effects of both bandwidth and center frequency. However, for a given bandwidth in Cams, the absolute widths (in Hz) of the noise bands increased with increasing center frequency. The thresholds may have been strongly influenced by the bandwidth in Hz because this bandwidth determines the average number of amplitude fluctuations per second and hence determines how many fluctuations occur within the stimulus duration. Figure 2 shows the geometric mean duration thresholds as a function of the bandwidth in Hz (log scale). Each center frequency is represented by a different symbol. A one-way repeated-measures ANOVA was conducted with factor bandwidth in Hz. A significant effect was found:
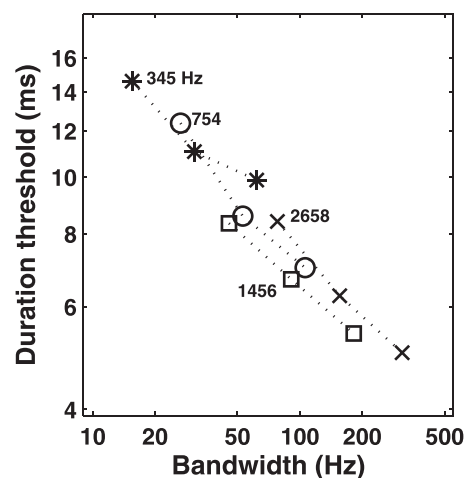

FIG. 2. Geometric mean duration thresholds across the 27 subjects of experiment 1 plotted as a function of bandwidth in Hz (log scale) with center frequency as parameter.

$[F(11, \ 286) = 39.8, \ p < 0.001]$. *Post hoc* comparisons showed that the duration threshold decreased significantly with increasing bandwidth whenever the two bandwidths that were compared differed by at least 65 Hz ($p < 0.05$).

Figure 2 shows that much of the variability in duration thresholds is accounted for by the width of the noise bands in Hz. The percentage of the variance in the data accounted for by the logarithm of bandwidth was 93%. This is consistent with the idea that the task was performed by detecting amplitude fluctuations in the noise and that the duration had to be sufficient for a detectable amplitude fluctuation to occur. However, bandwidth in Hz does not account for all of the variability in the data; the curves for the different center frequencies do not overlap completely.

III. EXPERIMENT 2

In experiment 2 the two bursts of sound in a given trial differed only in a short segment in their temporal center. One of the bursts was a sinusoid, and the other was the same sinusoid but with a small segment in the center replaced by a segment of noise (Taghipour *et al.*, 2014). These stimuli are relevant to the design of perceptual coders because the decision about tonality has to be made not for isolated frames but for frames embedded within other frames.

A. Method

1. Stimuli and apparatus

In every trial, the stimulus consisted of two 400-ms stimuli with a silent gap of 400 ms between them. Both stimuli were gated using a window function with raised-cosine ramps of 30 ms. One of the stimuli was a pure tone. The other was the same except for a short segment in the temporal center that was replaced by a narrowband noise of the same center frequency, generated in the same way as for experiment 1. To avoid discontinuities, cross-fading was used in the transition from tone to noise and back. The cross-fading windows had raised-cosine ramps. For overall noise durations up to 5 ms, the duration of each cross-fading ramp was 40% of the overall noise duration. For longer noise segments, the duration of the cross-fading ramps was kept constant at 2 ms. The noise segment that was actually picked (from a long narrowband noise burst) was determined based on the extent to which the waveforms of the tone and noise were similar in amplitude and phase within the two cross-fading ranges. The mean-squared difference between the time waveforms was used as a measure of similarity. As a result, the noise was faded in and out almost in-phase with the sinusoid [for more details, see Taghipour *et al.* (2014)]. This served to minimize spectral splatter. The estimated level of the stimuli at the eardrum was 65 dB SPL. This level was chosen to lead to a comfortable overall loudness, given the relatively long overall duration of the stimuli.

In the following, the overall duration of the middle noise segment (including half of the cross-fade ramps) will be referred to as the "noise duration," the transition sections as the "cross-fading ranges," and the resulting stimulus containing the noise segment as the "noisy stimulus." As for experiment 1, the noise bandwidth was 0.25, 0.5, or 1 Cams and the center frequencies were 345, 754, 1456, and 2658 Hz.

The experiment was carried out in a room that had a background noise level of 21 dBA. An artificial head (KEMAR, GRAS, Holte, Denmark) was used for calibration of levels. Stimuli were computed digitally with a sample frequency of 48 kHz and a resolution of 16 bits, converted to analog form using a Lawo DALLIS 941/83 digital-to-analog converter (Rastatt, Germany), and presented diotically via Sennheiser HD 650 headphones (Wedemark, Germany).

2. Procedure

A two-interval two-alternative forced-choice method was used. Subjects were asked to indicate which of the two bursts was a "pure tone." Feedback in the form of a green or red light was provided after each trial via a graphical user interface indicating a correct or an incorrect response, respectively. A hybrid staircase procedure was used. A run started with two "easy" trials using a rather long noise segment. Up to the third reversal point a 1-down/1-up method was used to achieve a rapid approach to the duration threshold. After that a 3-down/1-up method was used to estimate the 79% point on the psychometric function. The step size was 5 ms up to the second reversal point, then 2 ms up to the fourth reversal point, and then 1 ms until 10 reversals were obtained. The duration threshold for a run was calculated as the arithmetic mean duration at the last six reversal points. Two runs were obtained for each condition.

3. Subjects

Fifteen normal-hearing subjects participated. Audiometric thresholds were measured for frequencies up to 14 kHz using a software-based audiometer (SELFSCREENINGAUDIOMETER V1.32, HörTech GmbH, Oldenburg, Germany) and Sennheiser HDA 200 headphones. For all frequencies, the absolute thresholds for all subjects fell in the range -20 to $+20$ dB hearing level (ISO 389-7, 2005).

The "modified Thompson Tau test" and "Dixon's Q test" were used to check for outliers. Both tests revealed outliers for one or more conditions for four subjects, and all data for these four subjects were excluded from further analyses. Thus the final statistical analysis was based on thresholds for 11 subjects: eight males and three females. They were aged between 23 and 37 yr (mean 28 yr, median 26 yr).

4. Design

Each subject was tested in three sessions of about 30–35 min each, carried out on different days. Prior to the main experiment, subjects read a page of instructions. A training session of two runs with center frequencies 345 and 1456 Hz and bandwidths of 0.5 and 1 Cam, respectively, followed. This preparation/training phase took 20 min. During the main experiment, a break was given after every third run (i.e., after approximately 10–12 min). The conditions were presented in a random order.

J. Acoust. Soc. Am. **139** (5), May 2016

Taghipour *et al.* 2485

B. Results

Again, the variability in the thresholds was proportional to the mean. Thus geometric mean thresholds across subjects were calculated. Figure 3 shows the mean duration thresholds and their 95% confidence intervals (assuming that the data were normally distributed) on a logarithmic scale as a function of center frequency (logarithmically scaled abscissa). Shapiro–Wilks tests of normality showed that the data were normally distributed for nine conditions but deviated somewhat from normality for the other three conditions. A two-way repeated-measures ANOVA was conducted on the logarithm of the duration thresholds with factors bandwidth and center frequency. There were significant effects of bandwidth $[F(2, 20) = 153.6, p < 0.001]$ and center frequency $[F(3, 30) = 55.4, p < 0.001]$. There was no significant interaction; $[F(6, 60) = 0.9, p > 0.1]$.

LSD tests showed that duration thresholds decreased with increasing bandwidth (all pairs $p < 0.001$) and with increasing frequency up to 1456 Hz (all pairs $p < 0.01$). There was no significant difference between thresholds for the two highest frequencies ($p = 0.2$).

Figure 4 shows the geometric mean duration thresholds as a function of the bandwidth in Hz (log scale). Each center frequency is represented by a different symbol. A one-way repeated-measures ANOVA with factor bandwidth showed a significant effect; $[F(11, 110) = 43.0, p < 0.001]$. The logarithm of bandwidth accounted for 91% of the variability in the thresholds. LSD tests showed that duration thresholds decreased significantly with increasing bandwidth whenever the two bandwidths differed by at least 206 Hz ($p < 0.05$). As for experiment 1, the data showed that the duration thresholds were strongly influenced by bandwidth in Hz but that there was also an effect of center frequency. Specifically, for a given bandwidth in Hz duration thresholds were higher (worse) for the 2658-Hz center frequency than for the other center frequencies.
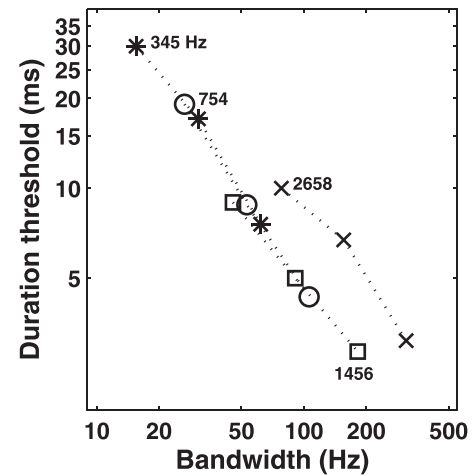
FIG. 4. Geometric mean duration thresholds across the 11 subjects of experiment 2 plotted as a function of bandwidth in Hz (log scale) with center frequency as parameter.

IV. DISCUSSION

A. Effects of spectral broadening

Decreasing the duration of the stimuli, as in experiment 1, would have led to a broadening of their spectra (Bos and de Boer, 1966; Moore, 2012). However, because the two stimuli that were being compared always had the same duration and were gated with the same window function, this spectral broadening would not be expected to provide a discrimination cue. If anything, the spectral broadening would make it more difficult to use any cues associated with differences in the spectra of the tone burst and the noise burst. In experiment 2, the noise bursts were embedded within longer tone bursts. In this situation, spectral broadening associated with the short noise burst embedded within the pure tone might have provided a cue for discrimination of the noisy stimulus from the pure tone. However, the characteristics of the individual noise bursts were chosen to minimize spectral "splatter" effects. Also, while the duration thresholds obtained in experiment 2 varied with bandwidth and center frequency in a similar way to the thresholds obtained in experiment 1, thresholds were generally higher in experiment 2; compare Figs. 1 and 3. This suggests that embedding the noise bursts within longer tone bursts made the task somewhat harder rather than providing an extra discrimination cue based on the detection of spectral splatter. Overall, it seems likely that performance of the two tasks was based on the temporal structure of the stimuli rather than on spectral cues.

B. Effects of frequency and bandwidth

The results for both experiments showed that duration thresholds decreased with increasing bandwidth. This is consistent with what has been found for another measure of temporal resolution, namely, the duration required to detect a gap in a band of noise (Eddins et al., 1992; Moore and Glasberg, 1988; Shailer and Moore, 1983). It seems likely that the tasks in experiments 1 and 2 were performed by the detection of amplitude fluctuations in the noise stimuli. One
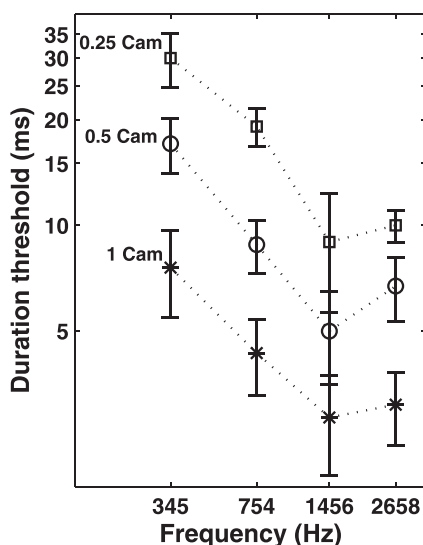

FIG. 3. Results of experiment 2: means and 95% confidence intervals of the duration thresholds across the 11 subjects are plotted as a function of frequency with noise bandwidth in Cams as parameter. Duration thresholds of the noise segments are shown in ms.

plausible hypothesis is that the duration threshold corresponds to a fixed number of envelope fluctuations, for example, a fixed number of envelope maxima. The number of envelope maxima per second increases with increasing bandwidth, and this could account for the dependence of the duration thresholds on bandwidth.

To test this hypothesis, the number of envelope fluctuations per second was determined empirically for the noise stimuli used in experiment 1. This was done by calculating the Hilbert envelope of a relatively long sample of the noise for each bandwidth and center frequency, locating the maxima, and calculating their number. Then the number of envelope maxima per second was multiplied by the mean duration threshold for the same center frequency and bandwidth. If the hypothesis is correct, the resulting number should be approximately constant and independent of the center frequency and bandwidth. The outcome is plotted in Fig. 5. In this figure, the mean number of envelope maxima in each stimulus at the duration threshold is plotted as a function of bandwidth. Each center frequency is represented by a different symbol.

Although the results for the different center frequencies cluster around a single function, it is clear that the function is not independent of bandwidth. Also, it is clear that the number of envelope maxima in the stimuli at the duration threshold is less than 1, especially for small bandwidths. In other words, the envelope needs to go through less than one "cycle" of fluctuation for the fluctuation to be detectable. The results suggest that duration thresholds are determined not only by how much the envelope fluctuates during the stimulus but also by the rapidity of the fluctuation; rapid fluctuations are harder to detect than slow fluctuations, consistent with temporal modulation transfer functions (Dau et al., 1997; Viemeister, 1979; Viemeister and Plack, 1993). It would be useful to conduct further experiments with an even wider range of center frequencies to assess the validity of this explanation.

A comparison of Figs. 1 and 3 shows that generally the thresholds were lower in experiment 1 and covered a smaller range. Isolated noise bursts seem easier to distinguish from tone bursts of the same duration than noise bursts embedded in a (longer) tone burst. This might be due to temporal uncertainty in the latter case or to forward and backward masking of the noise burst by the surrounding tone.

V. CONCLUSIONS

The results suggest that duration thresholds for discriminating a noise burst from a tone burst of the same duration and center frequency depend strongly on the envelope fluctuations in the noise stimulus. The thresholds also depend on the rapidity of the fluctuations. This information can be used in the design of more effective perceptual coders. The results also show that duration thresholds are slightly higher for noise bursts embedded within a longer tone burst than for noise bursts in isolation. This information can also be used in the design of more effective perceptual coders.
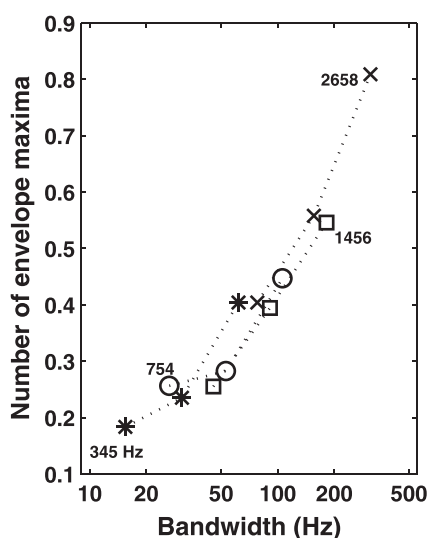
FIG. 5. Estimated mean number of envelope maxima in the noise bursts over the duration corresponding to the measured threshold, plotted as a function of bandwidth in Hz. Each symbol represents a different center frequency, as indicated in the figure (data from experiment 1).

Bos, C. E., and de Boer, E. (1966). "Masking and discrimination," J. Acoust. Soc. Am. 39, 708–715.

Bosi, M., Brandenburg, K., Quackenbush, S., Fielder, L., Akagiri, K., Fuchs, H., and Dietz, M. (1997). "ISO/IEC MPEG-2 advanced audio coding," J. Audio Eng. Soc. 45, 789–814.

Bosi, M., and Goldberg, R. E. (2003). *Introduction to Digital Audio Coding and Standards* (Kluwer Academic, Boston, MA), pp. 1–434.

Brandenburg, K., and Bosi, M. (1997). "Overview of MPEG audio: Current and future standards for low bit-rate audio coding," J. Audio Eng. Soc. 45, 4–21.

Brandenburg, K., and Stoll, G. (1994). "ISO/MPEG-1 audio: A generic standard for coding of high-quality digital audio," J. Audio Eng. Soc. 42, 780–792.

Chait, M., Poeppel, D., de Cheveigné, A., and Simon, J. Z. (2007). "Processing asymmetry of transitions between order and disorder in human auditory cortex," J. Neurosci. 27, 5207–5214.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. 102, 2906–2919.

Eddins, D. A., Hall, J. W., and Grose, J. H. (1992). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," J. Acoust. Soc. Am. 91, 1069–1077.

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. 47, 103–138.

Gockel, H., Moore, B. C. J., and Patterson, R. D. (2002). "Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression," J. Acoust. Soc. Am. 111, 2759–2770.

Hall, J. L. (1997). "Asymmetry of masking revisited: Generalization of masker and signal bandwidth," J. Acoust. Soc. Am. 101, 1023–1033.

Hellman, R. P. (1972). "Asymmetry of masking between noise and tone," Percept. Psychophys. 11, 241–246.

ISO 389-7 (2005). Acoustics - reference zero for the calibration of audiometric equipment. Part 7: reference threshold of hearing under free-field and diffuse-field listening conditions (International Organization for Standardization, Geneva, Switzerland).

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. 49, 467–477.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing*, 6th ed. (Brill, Leiden, The Netherlands), pp. 1–441.

Moore, B. C. J., and Glasberg, B. R. (**1988**). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears," J. Acoust. Soc. Am. **83**, 1093–1101.

Pollack, I. (**1968**). "Detection and relative discrimination of auditory 'jitter'," J. Acoust. Soc. Am. **43**, 308–315.

Shailer, M. J., and Moore, B. C. J. (**1983**). "Gap detection as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 467–473.

Sheft, S., and Yost, W. A. (**1990**). "Temporal integration in amplitude modulation detection," J. Acoust. Soc. Am. **88**, 796–805.

Stoll, G., and Brandenburg, K. (**1992**). "The ISO/MPEG-audio codec: A generic standard for coding of high quality digital audio," in *AES 129th Convention* (Vienna, Austria), paper 3336.

Stone, M. A., Moore, B. C. J., and Greenish, H. (**2008**). "Discrimination of envelope statistics reveals evidence of sub-clinical hearing damage in a noise-exposed population with 'normal' hearing thresholds," Int. J. Audiol. **47**, 737–750.

Taghipour, A., Edler, B., Amirpour, M., and Herre, J. (**2012**). "Thresholds for the discrimination of tonal and narrowband noise bursts," in *AES 133rd Convention* (San Francisco, CA), paper 8739.

Taghipour, A., Edler, B., Amirpour, M., and Herre, J. (**2013**)."Dependency of tonality perception on frequency, bandwidth and duration," Proc. Mtg. Acoust. **19**, 050040.

Taghipour, A., Guo, S., and Edler, B. (**2014**). "Thresholds for perception of rapid changes in tone bursts," in *Proceedings of 21st International Congress on Sound and Vibration (ICSV)*, paper R37.

Verhey, J. L. (**2002**). "Modeling the influence of inherent envelope fluctuations in simultaneous masking experiments," J. Acoust. Soc. Am. **111**, 1018–1025.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Viemeister, N. F., and Plack, C. J. (**1993**). "Time Analysis," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 116–154.