Quantifying Frequency Dependence of Auditory Search

Shuji Mori\(^1\) and Willy Wong\(^2\)

\(^1\) Department of Kinesiology, Faculty of Science, Tokyo Metropolitan University
\(^2\) Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada

Abstract  In the present study, we examined the frequency dependence of auditory search performance. Detection thresholds were measured for an 800-Hz target tone in a sequence of distractor tones (informational masking) as a function of frequency separation between the target and the distractor tones. The results showed that the thresholds decreased monotonically with frequency separation increasing. To further quantify the frequency dependence of auditory search performance, we applied a roex function (Patterson and Moore, 1986) to estimate a filter bandwidth for the threshold data. The estimated bandwidth was wider than that of an auditory filter by a factor of five (Leek et al., 1991). This result, together with some earlier results (Mori, 2003), demonstrates the effectiveness of the informational masking we used.

Introduction

In auditory search, a listener listens for a pre-specified target tone among non-target, distractor tones which are presented simultaneously or in temporal sequence with the target tone. Although the term ‘search’ has not been used in auditory research until recently (Asemi et al., 2003; Cusack and Carlyon, 2003), relevant studies have been conducted to investigate the human ability to find the target among the distractor tones. One such study is by Leek et al. (1991) where informational masking (Watson, 1987) was adopted to examine effects of tonal frequency and intensity on auditory search performance. In the informational masking, brief tones (distractors) of different frequencies were randomly presented in sequence and a target tone was inserted into the sequence. Leek et al. (1991) manipulated the target-distractor separation in frequency and in intensity, and measured the frequency discrimination thresholds for the target. The results showed that the thresholds decreased to an asymptote with increasing separation in frequency and increasing intensity difference between the target and the distractors.

Leek et al. (1991) modeled the effects of separation in terms of an attentional filter, which was calculated from the threshold results according to a roex\((p, r)\) function (Patterson and Moore, 1986; Patterson et al., 1982):

\[
W(g) = (1-r) (1+pg)e^{-pg} + r
\]  

where \(g\) is the relative difference between the target and the distractor tones (for the frequency separation \(g = |f-f_0|/f_0\) where \(f\) is the frequency of the distractor closest to the target frequency \(f_0\), and \(p\) and \(r\) are parameters determining the shape of the filter. In this function, \(4/p\) gives the ratio of equivalent rectangular bandwidth (ERB) of the estimated filter to the center (target) frequency. Leek et al. (1991) found that for the thresholds obtained by the target-distractor separations in frequency \(4/p\) was approximately 0.12 which is almost identical to the ERB ratio for an auditory filter calculated from thresholds under notched-noise masking (Patterson and Moore, 1986).

Although a close resemblance of the attentional filter to the auditory one is also reported in other studies (e.g., Dai et al., 1991; Schlauch and Hafter, 1991), it is at odds with the study by Leek et al. (1991) because informational masking is not energy masking like notched noise (which is typically used to measure the auditory filter) and has a much larger effect (Leek et al., 1991; Watson, 1987). There are two possible sources in the study by Leek et al. (1991) that may have contributed to the estimated filter being narrower than expected. Firstly, the target frequency (800–900 Hz) was always lower than the distractor tones, making it relatively easy to detect as compared with other studies where the target frequency was mixed with the distractor frequencies (Watson, 1987). Secondly, Leek et al. (1991) applied the roex function (Eq. (1)) to the frequency discrimination thresholds, although the roex function was designed for the detection thresholds (Patterson et al., 1982). In the present study, we measured the detection thresholds and calculated a filter for the target whose frequency was centered in the range of distractor frequencies (see Fig. 1) to examine the frequency dependence on auditory search performance.
Methods

Stimulus tones were generated by a sound generation system (Tucker and Davis Technologies, Experimenter II) controlled by a personal computer (IBM, PS/V model 2410), which also controlled timing and data collection. All tones were presented monaurally via a headphone (STAX, SR-Lambda PRO) to the listener’s left ear. A custom-made response box was connected to the computer and used to collect the listener’s responses and present warning signals and feedback on LED.

Target and distractors were sinusoidal tones of 46 msec in duration. A target was 800 Hz in frequency, and a set of eight distractors were separated in 1-semitone steps, as illustrated in Fig. 1. Four of the eight distractors were above 800 Hz, and the other four were below 800 Hz. The frequency separation between the target and the distractors was defined by the difference in semitones between the target and the nearest distractors in the upper and the lower groups of 4 tones. The intensity of the distractors was 75 dB SPL, while the target intensity was varied according to the listener’s response (see below).

In a given trial, a stimulus pattern consisted of a target and seven distractors that were presented in succession with no silent interval separating neighboring tones (Leek et al., 1991). For a given frequency separation, the seven distractors were composed of either one of the two distractors closest to the target frequency (just above or below 800 Hz) chosen randomly on every trial, and the 6 remaining tones. For example, the two distractors closest to the target for the 2-semitone separation are 713 and 897 Hz, one of which was chosen and presented with the other 6 tones and the target. Eight tones, one target and seven distractors, in the stimulus pattern were randomly ordered every trial, with the constraint that the target and the nearest distractor were not placed at either the beginning or the end of the stimulus pattern (Leek et al., 1991; Watson, 1987).

In the experiment, a trial began with a 500-ms warning signal on the LED display of the response box, followed by a stimulus pattern. Given a one-interval forced-choice task, the listeners were asked to report whether or not they heard a target tone in the presented pattern (note that a target was always present in the pattern). There was no time constraint for making a response, and following the response accuracy feedback was provided on the LED display for 1,000 ms. The threshold was measured by the QUEST staircase procedure (Watson and Pelli, 1983) which varied the intensity level of the stimulus pattern until they were familiarized with the stimuli. Thresholds were measured at 1-, 2-, 4-, 8-, and 10-semitone separations for all the participants, and most of them participated in the measurements at other separations as well.

Results

Figure 2 shows mean detection thresholds (in dB SPL) of seven listeners, as a function of frequency separation (in semitones). The thresholds decreased monotonically as the frequency separation was increased, from more than 85 dB at 1 semitone to near 75 dB at 10 semitone separation. Since the distractor intensity was 75 dB, further increases in frequency separation would result in the distractor tones being ineffective to mask the target tone, as the results of some listeners indicated.

Following Leek et al. (1991) we used the roex(\(p, r\)) function (Eq. (1)) to calculate the filter for our data. The integral of Eq. (1) represents the total ‘energy’ of the masker. The limits of integration were set equal to the upper and lower frequency limits of the masker. In the case of the masker lying above the target, the integral is of the form

\[
\int_{g}^{g'} W(g) dg = \left[-(1-r)p^{-1}(2+pg)e^{-pg} + rg\right]_{g}^{g'}
\]

where the upper limit \(g' = |f\times2^{1/2} - f_0|/f_0\) defines a frequency value of 3 semitones above \(f\) which is the distractor frequency closest to the target frequency \(f_0\) (800 Hz) for a given frequency separation (in semitones) \(s\), i.e. \(f = f_0 \times 2^{s/2}\). Similarly,
for the masker lying below the target,

$$\int_{h'}^{h} W(h) dh = -(1-r)p^{-1}(2+ph)e^{-ph} + rh f_0^2$$  \hspace{1cm} (3)

where $h = |f' - f_0|/f_0$ is the frequency difference between the target and the closest distractor frequency below the target $f' = f_0 \times 2^{-1/2}$, and $h' = |f' \times 2^{-1/4} - f_0|/f_0$ defines the lower bound of 3 semitones below $f'$. 

The difference in using the integral of roex($p$, $r$) function (Eqs. (2) and (3)) with the study of Leek et al. (1991) is that we have set the limit according to the width of the masker while Leek et al. (1991) took an upper value of infinity. To derive the expression governing the threshold, we note that the power of the target must exceed the masker for detectability. Thus the threshold in dB is given by ten times the logarithm of both the upper and lower masker ‘energy’ plus a constant to account for the efficiency of the detection process:

$$10 \log_{10} \left[ - (1-r)p^{-1}(2+pg)e^{-pg} + rh f_0^2 \right] + c$$

This is the equation which we fitted to the threshold data to obtain estimates of $p$ and $r$. The curve of the best fit as shown in Fig. 2 was obtained using $p = 6.56$, $r = 0.021$, and $c = 91.37$. Figure 3 shows the filter calculated from Eq. (1) with those values of $p$ and $r$. The ratio of ERB of the filter to the center frequency, given by $4/p$, is 0.61.

**Discussion**

The filter computed from our data was much wider than that obtained by Leek et al. (1991) and by studies using a noise masker (Patterson and Moore, 1986), exceeding a factor of 5 in terms of ERB (.63 vs .12). Our finding of the wide filter fits well with the notion that informational masking is much more effective than energy masking (Watson, 1987). Supportive evidence for the present results comes from our other study (Mori, 2003) which measured the signal detectability ($d'$) with stimuli almost identical to those used in the present study. The detectability increased gradually with increasing frequency separation and reached the highest value at only the 16-semitone separation. This corroborates with the present results in proving the effectiveness of the informational masking used in our studies.

By inspection of Eq. (4), we see that this is an equation with 3 free parameters. The experimental data shown in Fig. 2 suggests that the data roughly follows a straight line with only two degrees of freedom. In this case, the data will not likely permit a unique determination of 3 fitting parameters. As a result, we have also attempted to fit our data to the simpler roex($p$) filter (with one less adjustable parameter; Patterson and Moore (1986)) and the resulting fit was not appreciably poorer. We, however, keep our analysis with the roex($p$, $r$) filter in order to facilitate the comparison of the results with the results of Leek et al. (1991). Note that the theoretical curve shows a characteristic bend in its shape, and while the variability in the data does not permit a more precise determination of closeness of fit, this bend seems also to be found in the data points as well. We plan more work in a follow-up study to further explore the applicability of auditory...
filters to the attentional studies.

Acknowledgements This research was supported by a Grant-in-Aid for Scientific Research to Yoshitaka Nakajima from the Japan Society for the Promotion of Science (Grant No. 14101001).

References


Received: September 1, 2004
Accepted: October 4, 2004
Correspondence to: Shuji Mori, Department of Kinesiology, Faculty of Science, Tokyo Metropolitan University, 1–1 Minami-Ohsawa, Hachioji, Tokyo 192–0397, Japan
Phone: +81–426–77–2970
Fax: +81–426–77–2961
e-mail: smori@comp.metro-u.ac.jp