Auditory spectral discrimination and the localization of clicks in the sagittal plane

William Morris Hartmann
Department of Physics, Michigan State University, East Lansing, Michigan 48824
Brad Rakerd
Department of Audiology and Speech Sciences, Michigan State University, East Lansing, Michigan 48824

(Received 7 December 1992; revised 17 May 1993; accepted 15 June 1993)

Experiments show that the ability of human listeners to localize an impulsive sound in the medial sagittal plane (front, overhead, rear) deteriorates as the level of the sound increases. This *negative level effect* is strong for clicks but does not appear for broadband noise. It is conjectured that the negative level effect arises because the tonotopic excitation pattern is broadened for intense impulsive sounds. As a result, the spectral peaks and valleys, which are caused by anatomical filtering and which normally code for localization in the sagittal plane, are less recognizable. Filtered click discrimination experiments using headphones also show a negative level effect for clicks, but not for noise, and support this conjecture.

PACS numbers: 43.66.Qp, 43.66.Jh [HSC]

INTRODUCTION

There is now conclusive evidence that human listeners localize sounds in the medial sagittal plane (front-overhead-rear) on the basis of characteristic peaks and valleys in the spectrum. This spectral structure is the result of direction-dependent filtering by the listener's anatomy. Discrimination among sound source locations in the front portion of the sagittal plane is possible because of direction-dependent filtering by the pinnae. Because of their small size, the details of the pinnae only affect the frequency range above 7000 Hz (Roffler and Butler, 1968). Grosser anatomical features affect the spectral details over a broader range, down to frequencies as low as 200 Hz, and enable a listener to distinguish between sources in front and rear (Blauert, 1969; Searle *et al.*, 1975; Shaw, 1982; Asano *et al.*, 1990). Corresponding results have been seen in binaural recordings made with a KEMAR manikin (Kuhn, 1982).

Spectral cues permit the correct localization of a sound only if two conditions are met. First, the spectrum of the original source must be broadband and relatively flat in order to give the anatomical filtering a chance to manifest. If the signal is not broadband, the source cannot be localized correctly in the sagittal plane. In fact, when one tries to localize a narrow-band signal, a sine tone or a narrow band of noise, the apparent location is determined by the dominant frequency and not by the actual location of the source (Blauert, 1983; Rakerd and Hartmann, 1991). Second, the listener must have the ability to discriminate among the spectra corresponding to the different locations. A listener who is unable to resolve the details of the spectral profile established by anatomical filtering will not be able to localize sources in the sagittal plane.

The present article deals with the ability of listeners to localize clicks in the sagittal plane. A click signal satisfies the first of the requirements; it has a flat-broadband spectrum and can therefore convey the peak and valley information that mediates sagittal plane localization. Our particular interest in clicks arises because it may be possible to manipulate the listener's ability to resolve the spectral information in a click by varying the click level. An extensive history of temporal effects in masking, beginning with Zwicker (1965) and Elliott (1967), has been interpreted as indicating that auditory tuning is sharpened with increasing duration of the stimulus. A recent review has been given by Wright (1991). In particular, Bacon and Viemeister (1985a,b) studied the detection of brief signal targets in tonal maskers that were either continuous or gated. The difference in effective auditory filtering as measured with the two masking conditions grew substantially with an increasing stimulus level. Because clicks are the briefest stimuli of all, it seemed possible that this temporal effect might result in a decreasing ability to localize in the sagittal plane if the stimuli are impulsive and intense.

I. QUALIFYING EXPERIMENTS

Successful sagittal-plane localization requires an auditory system with a wide frequency range. Therefore, the listeners for the experiments were screened by two tests. First, a listener had to have thresholds within 25 dB of audiometric zero at ISO octave frequencies between 250 and 8000 Hz.[1] Second, listeners had to be able to localize sources of broadband noise that were in front, overhead, or at the rear.

A. Localization methods

The noise localization qualifying experiment employed an experimental geometry that was used for all the localization experiments discussed in this article. The listener was seated in the middle of an anechoic room (IAC 107840, with interior dimensions of 10×14×8-ft ceiling, i.e., 3×4.3×2.4 m). The six surfaces of this room are covered with 36-in. 0.91-m) foam wedges giving the room a cutoff frequency less than 100 Hz.

There were three loudspeakers [Minimus 3.5, consisting of a single 2.5-in. (6.4-cm) driver in a sealed box], one in front, one directly overhead, and one behind. The height of the listener's chair was adjusted to put the ears at the same level as the front and rear speakers. The speakers were all 48 in. (1.2 m) from the listener's ears. The three speakers had been selected from a lot of 17 on the basis of an automated transfer function comparison at 76 frequencies from 125 to 11 000 Hz. Those three speakers which were the most similar were selected. Of those speakers, the rms difference between the two least similar was less than 0.7 dB. The low-frequency limit, where the response was 10 dB down, was 125 Hz; the high-frequency limit was extended to 17.5 kHz by means of an equalizer. The stimulus was thermal noise (reverse-biased Zener selected for Gaussian probability density) band limited by the 125- to 17.5-kHz speaker response.

Listeners were asked to maintain their heads stationary. The back of the listener's chair was fitted with an adjustable L-shaped aluminum rod which was lowered so as to touch the top of the head. Using the sensation of the rod as a guide, the listener could minimize head motion. Also attached to the back of the chair was a fourth loudspeaker which created a background broadband noise (uncorrelated with the stimulus noise) with a level of 50 dB SPL. The background noise served as a masker against the possibility of weak switching transients. The above methods were used in all localization experiments in this article.

The qualifying experiment presented approximately 90 noise stimuli to a listener with front, rear, and overhead origins randomized. The listener's task was to identify the location of the source. The noises had a level of 65 dB SPL. They were turned on abruptly and turned off only after the listener made a response. The listener was aware of the number of sources and their locations, though only the front speaker was in view during the experiment.

B. Results

Initially, there were 12 listeners who passed the audiometric test. One of these individuals failed the noise-localization qualifying test. This listener correctly identified 36 presentations of the front source and 32 presentations of the rear, but on 24 presentations of the overhead source he gave a "rear" response 17 times (overall error rate of 18%). This listener did not participate in further tests.

The other listeners were clearly capable of localizing the noise. The overall percentage of errors made by the 11 listeners, in alphabetical order, was 0, 0, 1, 3, 0, 0, 0, 0, 3, 0, and 0. Summing over all listeners, there were no errors at all for the front source, five for the overhead source and two for the rear source. These 11 listeners qualified for the experiments to follow. Eight listeners (C, D, H, J, K, M, S, and T) were university students, four males and four females. Listener A was an older female. The coauthors, W and B, were also listeners.

II. LOCALIZATION OF CLICKS AND NOISE

The stimulus for the click localization experiment consisted of a train of eight clicks, 25 μs in duration and separated by 110 ms. The click train was presented by one of the three loudspeakers to make one experimental trial.

A. Methods

The principal variable in the experiment was the level of the clicks. There were six different levels. As measured with a peak-reading sound level meter (Larson–Davis model 800B), these levels were 98, 92, 86, 80, 74, and 68

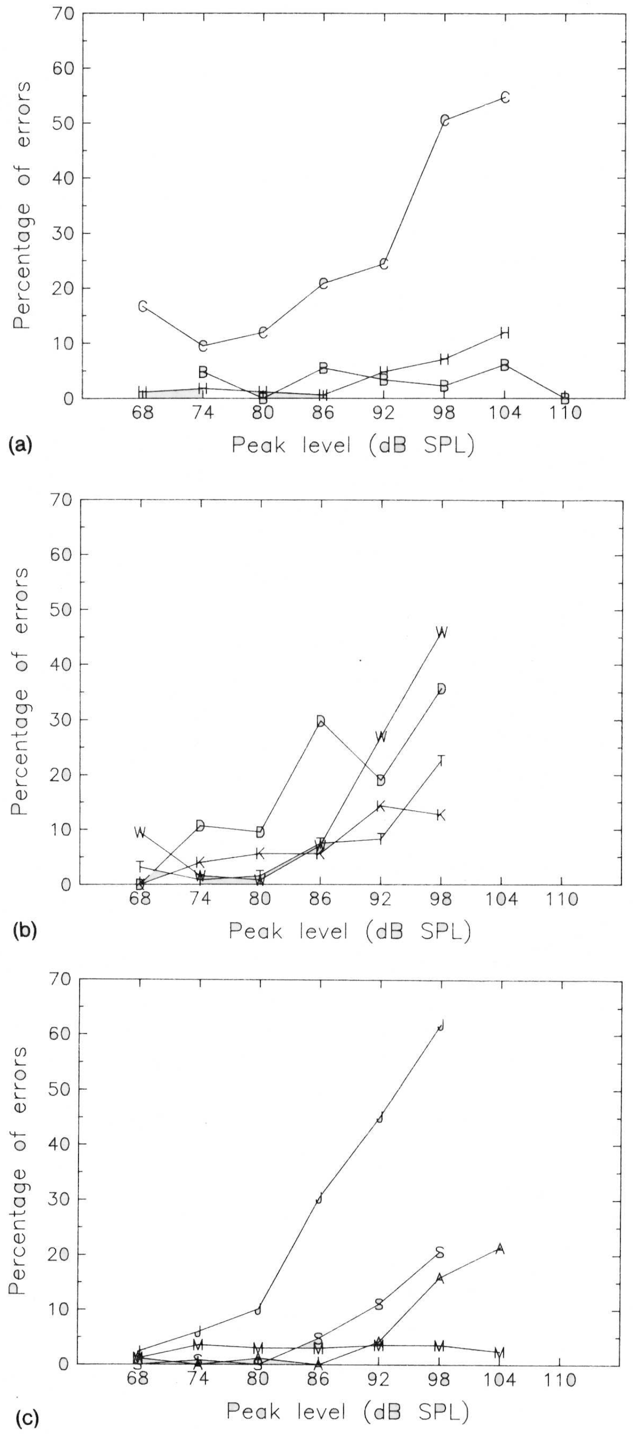

FIG. 1. (a)–(c) Errors in localizing a click in the sagittal plane, presented by front, overhead, and rear loudspeakers in an anechoic room. The percentage of errors is plotted for 11 listeners (identified by single letter symbols) as a function of the peak SPL of the click. Chance performance corresponds to 67% errors.

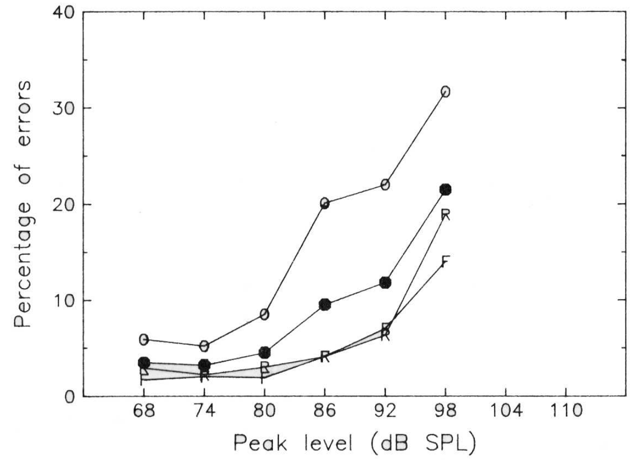

FIG. 2. Errors in localizing a click, averaged over the 11 listeners in Fig. 1 and expressed as a percentage. Separate plots show errors for presentation by front (F), overhead (O), and rear speakers (R). The solid symbols show the overall error and give the best average estimate of the level effect.

dB. The value of 68 dB corresponded to an average sensation level of 23 dB. A single experimental run combined three adjacent levels. Some runs were in the *high range* 98, 92, and 86 dB, some were in the *middle range* 86, 80, and 74 dB, and others were in the *low range* 80, 74, and 68 dB. Exceptional listeners were tested in a *very-high range* 104, 98, and 92 dB.[2] Each run consisted of 63 trials, 7 presentations from each of 3 speakers at each of 3 levels in random order. On each trial the listener had to choose whether the source was front, overhead, or rear. Listeners did three or four runs at each of the three standard level ranges.

B. Results

Figure 1(a)–(c) shows the percentage of incorrect responses for each listener as a function of the level, averaged over the three sources. It is evident that 9 of the 11 listeners were more successful at lower levels than at higher levels. Because an increase in signal level led to a decrease in performance, we shall refer to this result as a *negative level effect*.

Exceptions to the rule are listeners B and M, especially listener B for whom the error rate decreases with increasing level. To discover whether the error rate would go up at a higher level, listener B did an experiment in an *ultra-high* range 98, 104, and 110 dB. His performance did not decrease. In fact, he made no errors at all in this experiment.

Figure 2 shows the percentage of incorrect responses for each source, averaged over the 11 listeners. The figure shows that the overhead source caused more confusion than front and rear sources. Nevertheless, the errors for all sources increase with increasing level. The average over all sources and all listeners, given by solid points in Fig. 2, is our best estimate of the size of the negative level effect.

TABLE I. The percentage error in localizing broadband noise (front or overhead or rear) with different sound-pressure levels is given for ten listeners from the clicks experiment. Random guessing would lead to 67% error.

Nominal level	Listener									
(dB SPL)	A	B	C	J	H	K	M	S	T	W
44	0	0	0	0	0	0	0	0	0	19
64	0	0	0	1	0	0	0	1	0	0
84	3	0	0	44	1	0	0	2	0	1
90	0	1	1	43	1	0	0	0	1	3

C. Noise localization comparison

To discover whether the negative level effect seen with clicks could be duplicated with noise, we performed a noise-localization comparison experiment. The experiment was identical to the clicks experiment described above except that the train of eight clicks was replaced by 880 ms of broadband noise with abrupt onsets and offsets. For different experimental runs, different noise levels were used. Average rms levels were 44, 64, 84, and 90 dB SPL. For each run there were 18 trials with the noise level equal to the average level for the run; there were also 18 trials with the noise level 6 dB greater, and 18 trials with the noise level 6 dB less. Trials at different levels were randomized. Therefore, the minimum rms noise level was 38 dB and the maximum was 96 dB. Peak values of the noise were, of course, larger. Using the statistical methods of Hartmann and Pumplin (1988), we find that an ensemble average crest factor for our bandwidth and noise duration is 4.4. One can therefore expect peak levels that are 13 dB above the rms noise levels given above.

Ten of the 11 listeners from the clicks localization experiment did two runs of the noise localization experiment. The data are shown in Table I, giving the percentage of errors as a function of the average level. A strong negative level effect appears only for subject J. Upon repetition of the experiment several weeks later, the anomalous results for J persisted. A difficulty in interpreting the results of the noise localization experiment is that, unlike the clicks of the previous experiment, the intense noises were startling and unpleasant. It seems possible that some errors at the higher levels may have been the result of the aversive nature of the stimulus. Be that as it may, the data show that the negative level effect for localizing noise is not nearly as large or as common as the negative level effect for localizing clicks.

D. Discussion

The clicks localization experiment shows that, in most cases, the ability to localize clicks in the sagittal plane decreases as the click intensity increases. For some listeners the failure with increasing level is dramatic.

There are two kinds of explanation for the negative level effect. One of them regards it as a failure of the sensory system, which might be attributed to the broadening of excitation patterns in the auditory periphery, as suggested in the Introduction. The other regards it as a failure of postsensory interpretation. These explanations are described in some detail below.

1. A sensory explanation

In order for a listener to be able to localize a sound in the sagittal plane, the peripheral auditory system must encode the spectral peaks and valleys created by anatomical filtering. The information in the filtered sound resembles tone color information or vowel formant information.

The sensory explanation for the negative level effect supposes that the encoding of this information is tonotopic. Therefore, the peak and valley structure appears as a function of place along the basilar membrane and along tonotopic axes in higher centers too. However, the dynamic range of peripheral neurons is limited, which leads to saturation of the peripheral excitation pattern at high levels.

Physiological evidence in favor of this explanation comes from Sachs and Young (1979) and Young and Sachs (1979), who showed that the formant bands of vowel sounds are evident in a plot of eighth nerve firing rate versus neuron characteristic frequency for vowels presented at low levels. However, if the level is increased, the formant bands in the firing rate are obliterated because of saturation. It is possible that just such a saturation effect smears the spectral profile and prevents listeners from localizing high-level clicks.

The problem posed by the observation of Sachs and Young is that human listeners have no difficulty recognizing vowels at levels so high that they approach the threshold of pain. Similarly, the saturation effect should make it hard for listeners to localize high-level noise in the sagittal plane, a prediction which was not borne out experimentally for most of our subjects.

There are several escapes from the Sachs and Young problem for loud vowels. Proposed along with the dilemma itself was the conjecture that the synchronous firing rate, and not total firing rate, encodes the formant information. The synchronous firing rate is less susceptible to saturation. Alternatively, it may be that high-threshold low-spontaneous-rate fibers in the periphery encode vowel formants as the level is increased beyond the point where other fibers are saturated. Or perhaps small differences in firing rate on the auditory nerve are sharpened by inhibitory processes at somewhat higher levels of the periphery. (Winslow *et al.*, 1986). The parallel treatment of the localization problem would argue that synchrony, or the high-threshold fibers, or the pattern sharpening processes, do not operate so successfully in the case of signals that are as brief as our clicks (25 μs).

2. An interpretive explanation

Localizing a sound on the basis of anatomical filtering requires an interpretive operation by the central auditory system. Given a properly encoded set of spectral peaks and valleys, the system must decide whether this spectral structure is actually caused by anatomical filtering or whether the structure was present in the sound as it originated at the source. In fact, the central system fails to solve this puzzle in the case of narrow-band sounds, which is the basis of the narrow-band localization illusions found by Blauert (1969). But for wideband sounds, the system is remarkably successful. It apparently integrates place-based information over much of the tonotopic axis to find a self-consistent solution for source timbre and source location. It seems possible that one of the cues in this solution is the overall level, as a cue for the distance between the source and the listener. Conceivably the level of a sound becomes more important as a distance cue in an anechoic environment, such as ours, where there are no distance cues from wall reflections.

Basically, the interpretive explanation of the negative level effect says that judgements of front-overhead-rear locations are not decoupled from judgments of distance and externalization. Therefore, when the level of the click increases and the apparent source becomes closer to the head, listeners have a harder time determining the angular position of the source because sources that appear to be close together are more difficult to discriminate than sources that are well separated. This idea is consistent with the comment made by a number of the listeners when they listened to the intense clicks, "I can't tell whether the source is in front or overhead or in the rear; it's in my hair." The interpretive explanation says that in order to localize a sound in the sagittal plane a listener not only makes assumptions about the original spectral profile of the source, the listener also makes an assumption about the original intensity of the source. Promoting the use of the intensity cue may be the fact that the listeners knew the distance to the sources (see Mershon and King, 1975). The interpretive picture is no more successful than the sensory picture in explaining why there should be a negative level effect for clicks but not for noise.

III. DISCRIMINATION OF FILTERED CLICKS

To try to decide whether the sensory explanation or the interpretive explanation is better, we performed a headphones experiment where the listener's task was to discriminate between filtered clicks and clicks that had not been filtered. The filter transfer functions, when they were applied, had structure of the same scale as average anatomical filtering for external sources at different sagittal plane locations, but the headphone discrimination experiment eliminated the interpretive part of the localization task. Listeners were not required to localize the filtered clicks; they only had to distinguish them from unfiltered clicks.

A. Method

In the filtered-clicks discrimination experiment, clicks were presented via Sennheiser HD480 headphones. Listeners were tested individually in a sound-treated room. An experimental trial consisted of two intervals, one filtered, the other not filtered. Each interval consisted of a train of eight clicks, 25 μs in duration before filtering, and separated by 110 ms, just as in the localization experiment above. A gap of 300 ms separated the two intervals. The filtered clicks occurred in the first or second interval with

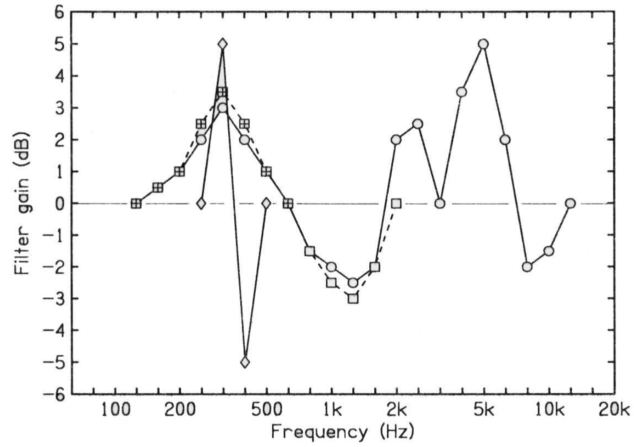

FIG. 3. Transfer functions for filters used in the discrimination experiments. Circles show the front-minus-rear [F−R] function taken from Blauert's book. Function [H&V] (squares) and [Hill] (+ symbols) are portions of function [F−R] with a scaling factor of 1.2. Transfer function [B&D] (diamonds) is not derived from HRTFs. It is shown here with extremes of ±5 dB. For those frequencies where no filter gain is shown the filter was flat (gain=0 dB). Experiments required listeners to discriminate filtered signals from unfiltered signals.

TABLE II. Listeners in the filtered clicks and filtered noise discrimination experiments. (Female listeners are starred.) Columns from left to right show: Threshold levels for unfiltered clicks and white noise, Scale factors for three broadband transfer functions, Extremes for the narrowband [B&D] transfer function.

	Thresholds (dB SPL)		Scale factors			
Listener	Click peak	White noise	[F-R]	[H&V]	[Hill]	Extremes [B&D]
A*	46	15	2	2	1.9	±5
B	50	21	1	1.6	3	±3
C*	50	23	2	2	2	±5
D*	47		2	2	3	
H	40	12	1	1	1.5	±3
J	43	13	1	0.8	1.5	±5
M*	45	12	2	2	1.9	±5
T	37	7	0.5	1	1.5	±3
W	45	16	0.5	1	1.2	±2
AVE	45	15				

equal probability. On each trial the listener decided whether the filtered clicks were on the first or second interval and indicated this choice by push buttons on a response box. After the response there was a feedback interval where the correct answer was given by pilot lamps on the response box. Listeners were expected to learn to recognize filtered clicks from this feedback.

There were 50 trials in an experimental run. In some early runs, it was evident that the listener had learned to recognize the filtered clicks within the first dozen trials causing performance to increase dramatically by midrun. Such runs were restarted in order to eliminate obvious learning effects. Because the listener could delay a response indefinitely there was no fixed duration for an experimental run. Done at maximum speed, a run lasted 2 min 35 s. The experimental data presented below were taken from the last four runs (200 trials) for a given condition.

1. Filtered clicks

A variety of different filters were used.[3] They were characterized by the gain (absolute value of the transfer function) in one-third-octave bands, as measured in dB. This measure is simply called "the transfer function" below. Filter transfer functions were variations of a standard pattern called "front minus rear [F−R]." The [F−R] pattern is a difference between average head-related transfer functions (HRTF) as measured with microphones in human ears. To determine the pattern [F−R], the HRTF with the source at the rear was subtracted from the HRTF with the source at the front (Blauert, 1983, p. 111). The [F−R] function is shown in Fig. 3. Also shown there are two variations, called "hill and valley" [H&V] and "hill." The [H&V] transfer functions consists of the lowest frequency peak and valley. The [Hill] transfer function consists of the lowest frequency peak alone. A fourth transfer function used in the experiment was called "bump and dip," [B&D]. The [B&D] pattern is the most rapidly varying transfer function possible with a one-third-octave filter. By contrast, transfer functions derived from [F−R] are slowly varying.

At the outset of the filtered clicks discrimination experiment, our intention was to use the transfer functions of Fig. 3 and to measure the ability of listeners to discriminate filtered clicks from unfiltered clicks at two different levels, 20 dB apart. The goal was to determine whether a level effect, either positive or negative, could be found in headphone listening. It was discovered, however, that individual listeners differed greatly in ability: whereas one listener might score almost perfectly at both levels, another might perform haphazardly for both levels. In neither case, of course, was anything learned about level dependence. The experiment was, therefore, tailored to individual listeners by scaling the filter transfer functions in order to get information on the effect of level. Scale factors are given in Table II for each listener. For [F−R], [H&V], and [Hill] functions the scale factors multiply the changes in decibels. For example, Fig. 3 shows that the first peak in [F−R] has a boost of +3 dB. When the scale factor is 2, in the condition called 2×[F−R], that peak is boosted to +6 dB. For the [B&D] transfer function, the characteristic is labeled by the extreme values, e.g., ±5 dB for the [B&D] shown in Fig. 3.

2. Levels

The overall gain of the filter was adjusted so that the total power was the same for filtered and unfiltered signals. During an experimental run, the peak level was fixed at one of two average peak levels, 92 or 72 dB SPL. The level of each train of eight clicks was then varied ±3 dB about the average value according to a random variable with a rectangular distribution. Roving the level prevented a listener from using overall loudness to discriminate the clicks. The rove of 6 dB also tended to discourage a listening strategy based upon the level in a single frequency band. The listener was expected to discriminate clicks on the basis of spectral shape.[4]

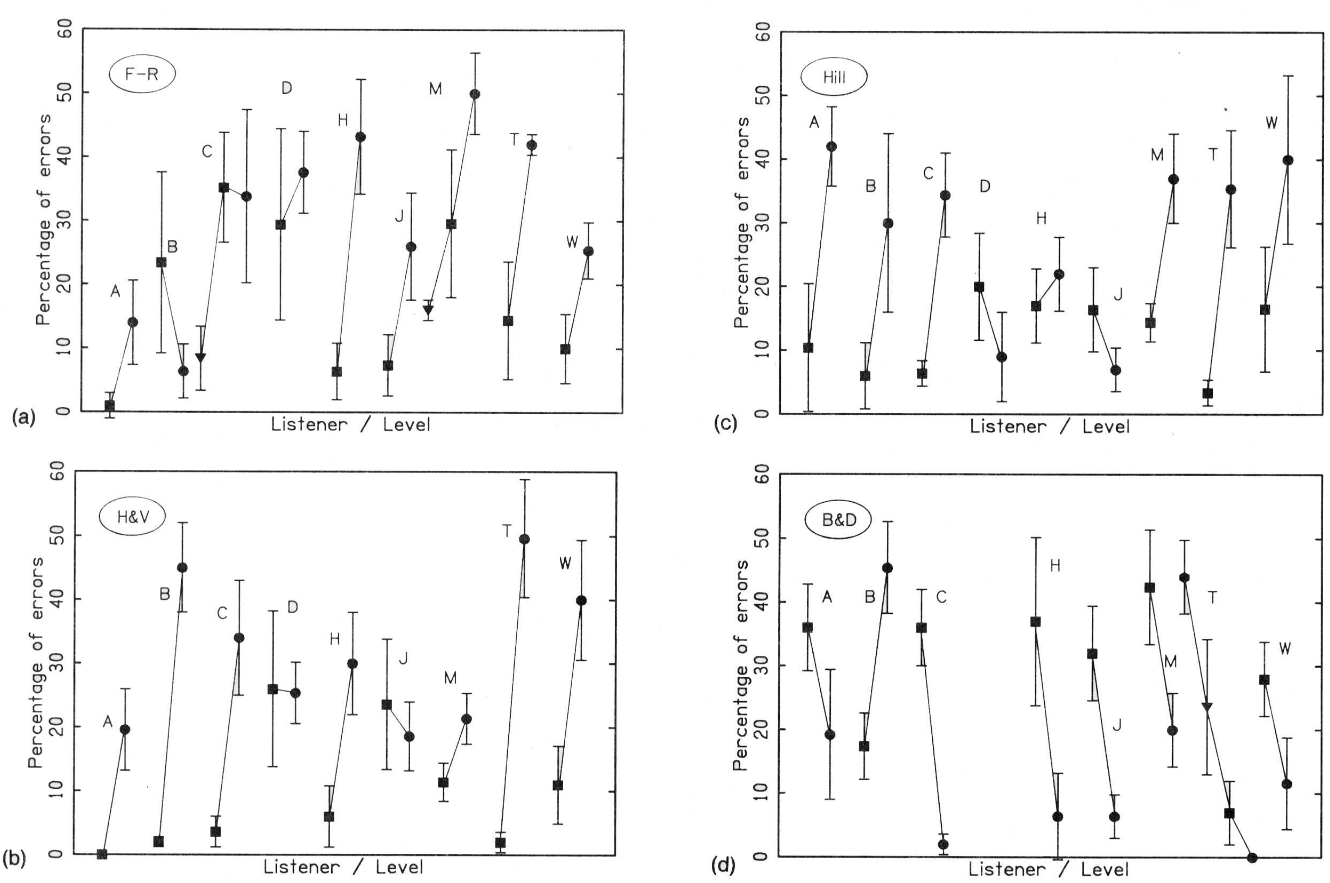

FIG. 4. Percentage of errors in discriminating between filtered and unfiltered clicks as a function of peak sound-pressure level of the click. Squares are for 72-dB-SPL clicks; circles are for 92-dB clicks. There are a few points at other levels, 62 dB (triangles) and 52 dB (pentagon). Lines connect data points for a single listener. There are nine listeners, indicated by letters of the alphabet. Each panel gives results for a different transfer function: in panel (a) [F−R], in panel (b) [H&V], in panel (c) [Hill], and in panel (d) [B&D].

B. Results

The effect of stimulus level on the ability to discriminate between filtered clicks and unfiltered clicks is shown in Fig. 4(a)–(d). There are four panels, one for each filter transfer function. The figure shows the percentage of error for each of the nine listeners for clicks presented at 72 dB SPL peak (squares) and for clicks at 92 dB SPL (circles). There are a few data points at still lower click levels too, as indicated in the figure caption. A line connects the data points for each listener. For the [F−R] transfer function [panel (a)] the slope of the line is positive; usually it is strongly positive. This increase in error rate with increasing intensity is a negative level effect, similar to the negative level effect seen in the sagittal plane localization experiments. There is one exception to this rule, listener B, who shows a positive level effect. Listener B was also anomalous in the clicks localization experiment.

The data for individual listeners in the [H&V] and [Hill] experiments resemble those in the [F−R] experiment. In these experiments, listener B shows a negative level effect like other listeners. There are small positive level effects for listeners D and J, but the sizes of the effects are small compared to the error bars. These weak exceptions do not seem important in comparison with the strong negative level effects observed for the other listeners.

Although the results for [H&V] and [Hill] experiments resemble results for the [F−R] experiment, we do not interpret the resemblance to mean that the [H&V] and [Hill] transfer functions are equivalent to the [F−R] transfer function. We do believe that listeners use similar discrimination strategies for [F−R], [H&V], and [Hill] most of the time. An exception is listener B who apparently uses the information at high frequency ([F−R]) differently from other listeners. Although there are individual exceptions, the strength of the negative level effect tends to be smallest for the [Hill] transfer function, where there is less structure in the transfer function.

The discrimination results for the [F−R] and [H&V] transfer functions support the sensory explanation for the negative level effect seen in localization, at least in general terms. The fact that listener B is exceptional on both localization and [F−R] headphone discrimination is further support. Among the other listeners, however, there is no tendency for those subjects who showed the largest level effect in localization also to show the largest level effect in discrimination. Correlation between rank-ordered performances in the two experiments is nonexistent.

The results for the [B&D] transfer function are completely different. The slope of the error rate as a function of click peak level is negative, which means that there is a *positive* level effect. There is one exception to this rule, the ever-anomalous listener B who shows a negative level effect.[5]

Important to the interpretation of the formal data were the answers given by listeners when asked about the cues used for discrimination. Somewhat surprisingly, there was almost no use of a location cue, even for the [F−R] transfer function. Listeners usually described the difference between clicks as "higher" or "lower." Probed further, they said that sometimes "high" and "low" referred to pitch, especially for the [B&D] function, but not always.

C. Discussion

The data show a convincing negative level effect for the [F−R] transfer function and for the two other broad transfer functions ([H&V] and [Hill]) that are portions of the [F−R] curve. By contrast, a *positive* level effect appeared for the spectrally sharp transfer function known as [B&D].

Physically, there is a distinction between the spectrally broad and the spectrally sharp clicks that can be observed in the temporal structure of the clicks, as seen on an oscilloscope. The spectrally broad clicks had ringing tails that were rapidly damped. The largest oscillation amplitudes were not small; they were as large as one-fourth the height of the initial click. However, successive oscillation peaks in the ringing response were irregularly spaced. The spectrally sharp [B&D] clicks had tails with longer ring times, but the largest oscillation amplitudes were a hundred times smaller than the height of the initial click. Although the oscillations for the [B&D] clicks were small, the peaks were regularly spaced in time. One could imagine that the regular oscillations would excite a nervous system element that is sensitive to periodicity, if the oscillations had sufficient intensity.

Informal comments from the listeners were consistent with the time-domain physical description in that the spectrally sharp [B&D] clicks were described as hollow or drumlike, at least at the high level. Comments on the color of the clicks with broad spectral structure did not refer to this aftersound.

Our interpretation of the level effect is that when the spectral structure is broadly spread in frequency there is only aperiodic ringing, but the click itself acquires a characteristic timbre because a sizeable range in frequency is affected by the filtering. The timbre depends upon the tonotopic encoding of the spectral structure at the instant of the click. This tonotopic encoding is vulnerable to saturation effects at high levels which accounts for the negative level effect.

By contrast, when the spectral structure is sharp, the effects on the tonotopic encoding of the click are localized in a small region and are, therefore, not very important. However, the almost-periodic ringing creates an aftersound that enables listeners to recognize filtered clicks, so long as the level of the entire stimulus is high enough that the relatively low-level ringing can be heard. When spectrally narrow clicks are heard at low level, the ringing may be below threshold (or below the forward-masked threshold given the body of the click as the masker) which accounts for the positive level effect seen with the [B&D] stimulus.

D. Caveat

It should be noted that the discrimination experiment enabled us to ignore many details in creating the headphone sound. We did not attempt to simulate individual transfer functions for sources in front, overhead, and rear positions. These are highly variable among listeners. Nor did we need the unfiltered clicks to be optimally flat. Our concern with the headphone response was limited to ensuring that it had no abrupt peaks or valleys in the frequency region of the filtering.

The modest goal of the discrimination experiments was to determine whether listeners were sensitive to a spectral *change* comparable to a major difference in sagittal plane angle. The average [F−R] characteristic was chosen as a comparable difference. An advantage of the [F−R] characteristic is that there is spectral structure at frequencies as low as 300–2000 Hz. At such frequencies, one avoids the sensitive headphone problems that occur at frequencies above 7 kHz, typical of pinna spectral cues that mediate the perception of elevation. Because of the methodological differences, the discrimination experiments posed quite a different task from the localization experiments. Common to both tasks was the need to resolve spectral structure in broadband impulsive stimuli.

IV. DISCRIMINATION OF FILTERED NOISE

The sensory explanation for the negative level effect observed in the localization of clicks asserts that the peripheral auditory system copes less well with a high-level impulsive sound like a click than with a more continuous sound like noise. To test this idea, we supplemented the filtered click experiments with filtered noise discrimination experiments.

A. Methods

1. Procedure

The same filter settings were used for both click and noise experiments. In fact, the two experiments were identical except that in the original signal source, thermal white noise was substituted for the train of clicks. Therefore, the two noise intervals were 880 ms in duration and were separated by 300 ms. The listener's task was to distinguish the filtered noise from the unfiltered noise. Because the original noise source was white, the long-term spectra of noise and clicks were identical; only the phase spectra were different.

The noise discrimination experiment was done together with the click discrimination experiment. Generally, click runs were done first to find an appropriate filter setting, and then noise experiment runs were mingled with further click runs. In all, listeners did three noise runs, 150 trials.

2. Level

Because the noise discrimination experiment was a control for click discrimination, it made sense to run the

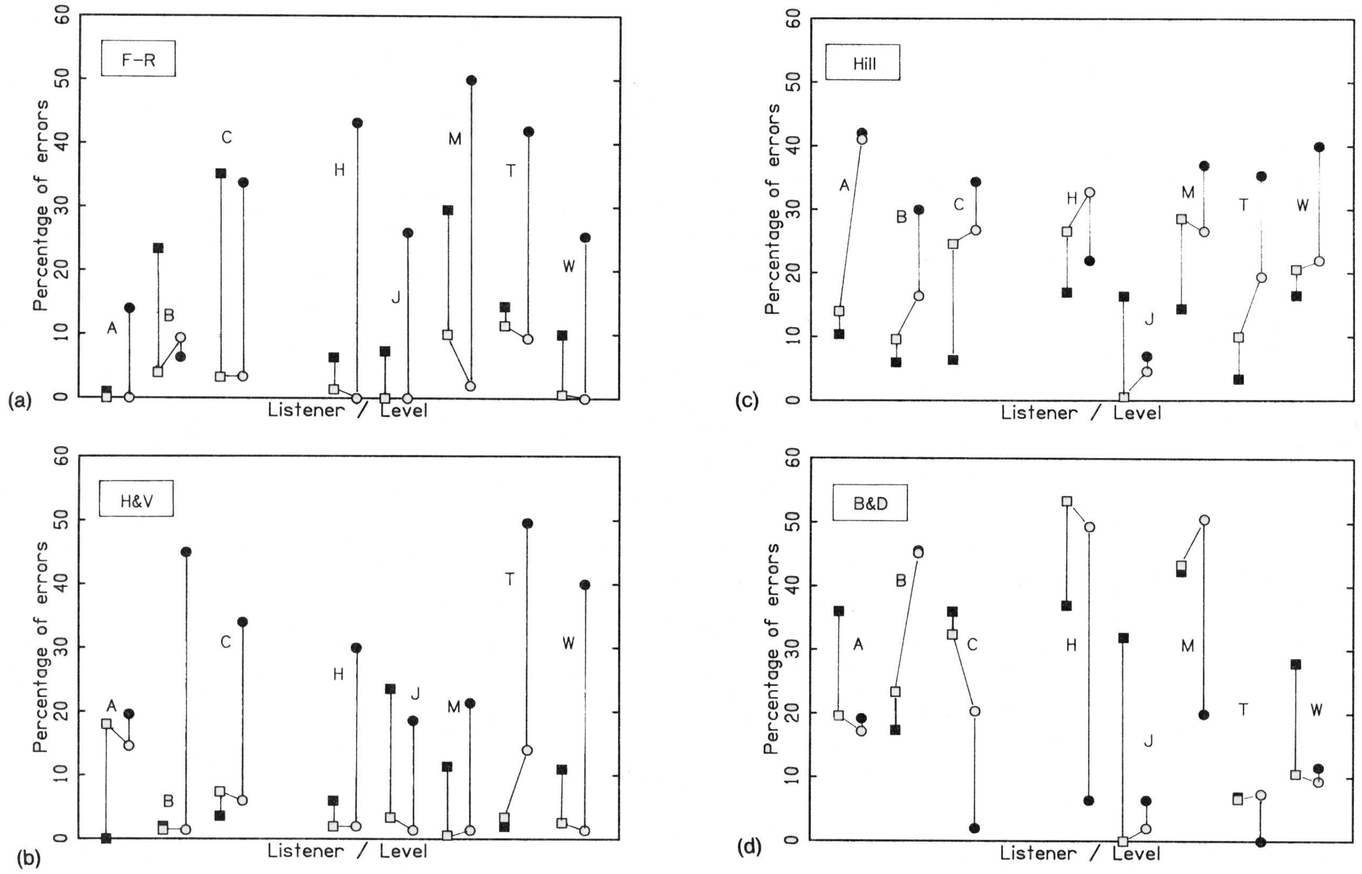

FIG. 5. Percentage of errors in discriminating between filtered and unfiltered stimuli. Open symbols show the result for noise; closed symbols are for clicks, as in Fig. 4. Squares are for low level (72 dB SPL clicks and 49 dB noise). Circles are for high level (92 dB SPL clicks and 69 dB noise). Lines connect data points for a single listener. There are four different filter transfer functions, in panel (a) [F—R]: in panel (b) [H&V], in panel (c) [Hill], and in panel (d) [B&D].

two experiments at comparable levels. However, there is no obvious way to equate the levels of a click and a continuous noise. Our approach was to choose a reference intensity for the continuous noise such that the energy integrated over 3 ms was equal to the energy in the higher level (92-dB) click. To do this we worked with signal power. We made a train of high-level clicks, separated by 3 ms, and measured the intensity at the headphones with a flat-plate coupler and a C-weighted sound level meter. The meter read 69 dB SPL. The high-level noise was then set to give the same reading. If the auditory system has an integration time of 3 ms, then this procedure should lead to equal sensation levels for clicks and noise. In fact, the procedure did not lead to equal sensation levels for clicks and noise, though they were not greatly different. As shown in Table II, the average threshold for clicks (eight clicks separated by 110 ms) was a peak reading of 45 dB SPL. Therefore, the high-level clicks at 92-dB peak were 47 dB above threshold. The table also shows that the average threshold for noise was 15 dB SPL. Therefore, the high-level noise at 69 dB was 54 dB above threshold. Sensation levels of 47 and 54 dB are not greatly different.

B. Results

The percentage of errors in noise discrimination is given in Fig. 5(a)–(d) (open symbols) together with a repeat of the data for the clicks (filled symbols). The comparison of noise and click data show clearly that for transfer functions [F—R] and [H&V] the error rate for noise is lower than the rate for clicks. The filled symbols lie above the equivalent open symbols in 28 out of 32 instances, and there is only one instance in which an open symbol lies appreciably above the closed symbol. The difference is particularly marked for the [F—R] transfer function.

For the [Hill] and [B&D] transfer functions the comparison is less clear because the absolute values of error rates for noise are elevated. For [Hill] the error rates for noise tend to fall between the error rates for low- and high-level clicks.

The connecting lines between open symbols in Fig. 5 show the nature of the level effects for noise. Overall, level effects appear to be positive about as often as they are negative. A possible exception is the [Hill] condition where there may be a negative level effect. There is, however, no level effect for noise that approaches the importance of the level effect seen for clicks.

C. Discussion

The error data show that listeners were very successful in discriminating noise passed by [F—R] and [H&V] transfer functions. We interpret this result to mean that listeners perform an analysis of the spectral profile of the noise and can readily distinguish between noise timbres by using the broadband structure of the transfer functions over wide regions of the frequency axis. The considerably larger error rate when transfer functions with narrower structure ([Hill] and [B&D] are used may reflect only the smaller available spread on the frequency axis, which ought to make the task harder. Alternatively, listeners may adopt a different strategy in the case of these narrower transfer functions. They may listen for the ringing of the filter, which appears as a "howling" sound, especially for the [B&D] filter.

The lack of any consistent level effect for the noise experiment distinguishes the perception of filtered noise from the perception of filtered clicks. This result tends to support the proposition that the negative level effect seen for broadband clicks is a direct result of the impulsive nature of the clicks and indicative of a peculiar problem that the auditory system has with such impulsive sounds. However, it should be noted that our procedure selected transfer function scale factors to optimize level effects (negative or positive) for clicks. The same scale factors were then used for the noise experiments, and it is possible that they missed finding level effects for noise because they were not optimized for noise. Based upon the many different transfer functions and levels actually explored in the course of doing these experiments, we consider this scenario to be unlikely, but it is anyhow a theoretical possibility.

V. CONCLUSION

Listening experiments in an anechoic room showed that the ability to determine whether a click originates from a source that is in front or overhead or behind decreases as the intensity of the click increases. This effect was called the "negative level effect." No comparably strong effect was observed for white noise, which has the same long-term spectrum as the click. Apparently, the impulsive nature of the click is responsible for the effect.

The negative level effect appears to be peculiar to sagittal plane localization. In azimuthal plane localization, or in lateralization, performance tends to improve with increasing level. For example, Dye and Hafter (1984) found that thresholds for interaural time differences between pulses decrease with increasing pulse level, and they cite five other studies in which the interaural sensitivity increased with increasing level.

It was conjectured that the negative level effect arises because the peripheral auditory system fails to resolve the spectral details of clicks, as filtered by the direction-dependent anatomical transfer function. To test this idea we measured the ability of listeners to distinguish between filtered and unfiltered clicks, as heard through headphones.

The headphone experiments used four different transfer functions for filtered clicks. Three of them had broadband structure (broader than a critical band) resembling average anatomical transfer functions. Experiments with these broadband transfer functions showed a negative level effect, supporting the idea that a deficit in peripheral spectral resolution is responsible for the negative level effect seen in localization. The fourth transfer function had a narrow-band structure, which led to a positive level effect. It seems likely that the positive level effect arises when listeners recognize filtered clicks on the basis of a ringing response which imparts a tonal aftersound to the click.

Evidence in favor of a tonotopic explanation for the perception of broadband filtering, as opposed to narrow-band filtering, came from the comparison of filtered click discrimination with filtered noise discrimination. Broadband filtered noise was easier to recognize than broadband filtered clicks. Apparently, the persistence in time allows the auditory system to make a more precise comparison of excitation at different places along the tonotopic coordinate in the case of noise, leading to better discrimination.

For narrow-band filtering a different result occurred. Narrow-band filtered noise was no easier to recognize than narrow-band filtered clicks. The explanation for this result is that the ringing response of the filters may be partly masked by the continuous noise.

Conclusions similar to ours, distinguishing between broad and narrow resonances, have been reached by audio engineers, who have studied the audibility of single resonances as an element of loudspeaker design. Toole and Olive (1988) found that with anechoic conditions or headphone listening, a broad resonance was much better revealed by continuous noise than by an impulse. By contrast, a narrow resonance was equally revealed by continuous noise and by an impulse. They cite unpublished work by Moulana (1975) showing that a broad resonance with $Q < 10$ can be more easily heard with continuous noise. For a sharp resonance with $Q > 50$, however, it does not matter whether one uses noise or clicks.

The advantage of persistence in time for the detection of broadband resonances was shown in an experiment by Toole and Olive where the difference between continuous noise and clicks largely disappeared when the experiment was moved from an anechoic room to an ordinary room, where reflections repeated the clicks. This result was not tested in our experiments, which were confined to anechoic conditions or to headphone listening. Informally, however, the effect was evident to experimenters monitoring the experiments via loudspeakers in an ordinary room. Moore *et al.* (1989) gave additional evidence for a duration effect, suggesting that resonances with Qs between 2 and 8 are better observed with 200-ms noise bands than with 100-ms noise bands.

In summary, the principal new effect described in this paper is the decreasing ability to resolve the spectral details of filtered impulsive sounds as the level of the sounds is increased. This deficit can be seen in headphone experiments on click discrimination, and it also is manifest as a decreasing ability to localize clicks in the medial sagittal plane with increasing level. Details of this effect are important: The stimuli must be impulsive; there is no comparable effect for continuous noise. The spectral structure must be broadband; the sign of the effect is reversed if the structure is narrow band. The localization effect was observed in an anechoic environment. It is not known whether the effect would persist in an ordinary room, but it seems likely that a change in the room would affect the data in detail.

The results of the click and noise experiments are relevant to auditory system modeling and to the interpretation of neural coding of acoustic space (Brugge *et al.*, 1992). If one assumes that the peripheral system is a bank of linear filters with center frequencies coded by place, and that the filter parameters are time independent, and that the perceived spectrum is a long-term average of the filter outputs, then one might expect the perceived spectrum to be the same for clicks and for noise, because these two stimuli have the same average power spectra. However, the localization experiments and discrimination experiments show that the spectral details are perceived better for noise than for clicks. Therefore, one or more of the assumptions needs to be changed. As suggested in the Introduction, one can begin with the assumption of stationary filter parameters. For example, if auditory bandwidths reflect a competition between excitation, which broadens with increasing level, and lateral inhibition (a nonlinear band-narrowing process) and if the operation of inhibition is delayed compared to excitation, then the details of noise should be resolved better than the details of intense clicks, in agreement with experiment. Alternatively, a combination of saturation of excitation peaks and adaptation in excitation valleys could mean that at high levels tonotopic details are flattened for brief stimuli but are recovered for extended stimuli. Once one admits nonlinearities into the auditory model, it becomes relatively easy to find principles that can explain our observation that the auditory system copes less well with intense clicks than with noise.

ACKNOWLEDGMENTS

Ms. Maggie Whearty and Mr. T. J. Vander Velde measured hearing thresholds for the listeners in this experiment. Mr. Mitch Milton from Applied Research and Technology provided technical information on the Intelligent Equalizer.

---

[1] Thresholds were measured using a standard audiometer. Listeners B and W showed some high-frequency hearing loss, with 8000-Hz thresholds at 25 dB SPL in the better ear.

[2] Although these peak levels may seem high, the clicks were brief and did not sound loud to the listeners.

[3] The filter used to create transfer functions was an intelligent equalizer from Applied Research and Technology. This is a bank of 32 two-pole filters with 1/3-octave spacing between the bands. The "intelligence" implements an algorithm that readjusts the gains of each filter to minimize cross-band interaction. If, for example, the user boosts a single band then the algorithm automatically reduces gains in neighboring bands to maintain a maximally flat amplitude transfer function outside the boosted band.

[4] The rove of 6 dB is small in comparison with roves used in profile analysis studies, as large as 20 dB in noise-band-slope discrimination (Versfeld, 1992) and as large as 40 dB for line spectra (Kidd *et al.*, 1986). Large roves are awkward in our experiment, designed to measure a level effect. What is most important, however, is that the conditions of the headphones experiment resembled the conditions of the localization experiment.

[5] The negative level effect for B was so surprising that we repeated the experiment a week later with a bump and dip of ±4 dB instead of ±3 dB. The error rates for -40-dB clicks and -20-dB clicks both improved, as would be expected, but the negative level effect persisted: 5% error at 72 dB and 27% error at 92 dB.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). "Role of spectral cues in median plane localization," J. Acoust. Soc. Am. **88**, 159–168.

Bacon, S. P., and Viemeister, N. F. (**1985a**). "Simultaneous masking by gated and continuous sinusoidal maskers," J. Acoust. Soc. Am. **78**, 1220–1230.

Bacon, S. P., and Viemeister, N. F. (**1985b**). "The temporal course of simultaneous tone-on-tone masking," J. Acoust. Soc. Am. **78**, 1231–1235.

Blauert, J. (**1969**). "Sound localization in the median plane," Acustica **22**, 205–213.

Blauert, J. (**1983**). *Spatial Hearing* (MIT, Cambridge, MA).

Brugge, J. F., Chan, J. C. K, Hind, J. E., Musicant, A. D., Poon, P. W. F., and Reale, R. A. (**1992**). "Neural coding of virtual acoustic space," J. Acoust. Soc. Am. **92**, 2333 (A).

Dye, R. H., and Hafter, E. R. (**1984**). "The effects of intensity on the detection of interaural differences of time in high-frequency trains of clicks," J. Acoust. Soc. Am. **75**, 1593–1598.

Elliott, L. L. (**1967**). "Changes in simultaneous masked thresholds of brief tones," J. Acoust. Soc. Am. **38**, 738–746.

Hartmann, W. M., and Pumplin, J. (**1988**). "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.

Kidd, G., Mason, C. R., and Green D. M. (**1986**). "Auditory profile analysis of irregular sound spectra," J. Acoust. Soc. Am. **79**, 1045–1053.

Kuhn, G. F. (**1982**). "Towards a model for sound localization," in *Localization of Sound: Theory and Applications*, edited by R. W. Gatehouse (Amphora, Groton, CT).

Mershon, D. H., and King, L. E. (**1975**). "Intensity and reverberation as factors in the auditory perception of egocentric distance," Percept. Psychophys. **18**, 409–415.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 9 kHz," J. Acoust. Soc. Am. **85**, 820–836.

Moulana, K. (**1975**). "Tonal coloration caused by a single resonance," thesis, University of Surrey, U.K.

Rakerd, B., and Hartmann, W. M. (**1991**). "Localization of narrow-band sounds in the medial sagittal plane," J. Acoust. Soc. Am. **90**, 2266 (A).

Roffler, S. K., and Butler, R. A. (**1968**). "Factors that influence the localization of sound in the vertical plane," J. Acoust. Soc. Am. **43**, 1255–1259.

Sachs, M. B., and Young, E. D. (**1979**). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," J. Acoust. Soc. Am. **66**, 470–479.

Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (**1975**). "Binaural pinna disparity: another auditory localization cue," J. Acoust. Soc. Am. **57**, 448–455.

Shaw, E. A. G. (**1982**). "External ear response and sound localization," in *Localization of Sound: Theory and Applications*, edited by R. W. Gatehouse (Amphora, Groton, CT).

Toole, F. E., and Olive, S. E. (**1988**). "The modification of timbre by resonances: Perception and measurement," J. Audio Eng. Soc. **36**, 122–142.

Versfeld, N. J. (**1992**). "On the auditory discrimination of spectral shape," thesis, Technical University of Eindhoven, The Netherlands.

Winslow, R. L., Barta, P. E., and Sachs, M. B. (**1986**). "Rate coding in the auditory nerve," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 212–224.

Wright, B. A. (**1991**). "Contributions of stimulus bandwidth, the reference condition, and individual differences to the auditory temporal effect: Evidence for within- and across-channel mechanisms," thesis, University of Texas, Austin, TX.

Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403.

Zwicker, E. (**1965**). "Temporal effects in simultaneous masking by white-noise bursts," J. Acoust. Soc. Am. **37**, 653–663.