Identification and localization of sound sources in the median sagittal plane

Brad Rakerd[a)]
Department of Audiology and Speech Sciences, Michigan State University, East Lansing, Michigan 48824

William M. Hartmann[b)] and Timothy L. McCaskey[c)]
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 7 December 1998; revised 16 June 1999; accepted 10 July 1999)

The ability of human listeners to identify broadband noises differing in spectral structure was studied for multiple sound-source locations in the median sagittal plane. The purpose of the study was to understand how sound identification is affected by spectral variations caused by directionally dependent head-related transfer functions. It was found that listeners could accurately identify noises with different spectral peaks and valleys when the source location was fixed. Listeners could also identify noises when the source location was roved in the median sagittal plane when the relevant spectral features were at low frequency. Listeners failed to identify noises with roved location when the spectral structure was at high frequency, presumably because the spectral structure was confused with the spectral variations caused by different locations. Parallel experiments on sound localization showed that listeners can localize noises that they cannot identify. The combination of identification and localization experiments leads to the conclusion that listeners cannot compensate for directionally dependent filtering by their own heads when they try to identify sounds. [S0001-4966(99)00111-3]

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

INTRODUCTION

The localization of sources of sound in the median sagittal plane (MSP) presents an unusual problem to a listener. The MSP includes the points directly in front, directly behind, and overhead. Therefore, it is symmetrical with respect to the two ears, and binaural differences—usually of paramount importance for sound localization—are minimal or absent. Nevertheless, most normally hearing individuals can successfully localize sounds in this plane. To do so, listeners use the fact that the external ears and head serve as an acoustical filter with a frequency response that depends on the direction of the sound source. Such a frequency response can be measured with probe microphones in the ear canals (Shaw, 1966; Hebrank and Wright, 1974; Wightman and Kistler, 1989a, b). The response to a source in front has a peak and valley structure that favors 4000 Hz compared to 1000 Hz. The response to a source in back has a response with the opposite peak and valley structure. A source overhead leads to a response curve with a peak near 8000 Hz. It is individual characteristic spectral features like these that are thought to enable listeners to localize in the absence of binaural difference cues (Butler and Belundiuk, 1977).

A logical problem with the process of localization on the basis of spectral features is that different real-world sources emit sounds with very different spectra. Spectral structure serves to identify sounds, for example to distinguish between different vowels. *A priori*, a listener does not know whether a particular spectral structure is caused by location-dependent filtering or whether it is intrinsic to the source itself (Durlach and Colburn, 1978). A sound with a spectral bump at 8000 Hz may not come from above; instead it may come from a source that happens to have considerable power at 8000 Hz. Confusions of this kind were dramatically demonstrated by Blauert (1969–70; also see Middlebrooks, 1992), who studied the perception of *narrow bands* of noise using loudspeakers that were in front, in back, and overhead. The perceived location of sources in the experiment was shown to be unrelated to the actual position of the loudspeakers. Instead, the perceived location of each noise band was determined by its frequency, whatever the true location of the source.

Blauert's experiment showed that for narrow-band signals in the MSP, the confusion between the spectral identity of a source and the location of the source is essentially complete. One can imagine that confusion would extend to broadband sounds as well, though the confusion might be less complete. Plenge and Brunschen (1971) conjectured that listeners would localize a broadband sound more successfully if they were familiar with it. With a familiar sound, the intrinsic spectrum might be regarded as known. Then, additional spectral structure could be interpreted unambiguously as a location cue. Plenge and Brunschen used speech fragments from known and unknown talkers. During a training phase, speech from the known talkers was presented from each of five loudspeakers in the upper-half MSP. In the test phase, speech from unknown talkers was sometimes substituted. Plenge and Brunschen found that in the test phase, correct localization dropped from near 90% for known talkers to much less than 50% for unknown talkers, thus supporting their conjecture.

The present article also studies the relationship between the identification of a sound and the localization of its source. It begins by asking the reverse of Plenge and Brun-

[a)]Electronic mail: rakerd@pilot.msu.edu
[b)]Electronic mail: hartmann@pa.msu.edu
[c)]Electronic mail: mccaskel@pilot.msu.edu

schen's question: Is it more difficult to identify a sound if its location in the MSP is uncertain?

I. EXPERIMENT 1—IDENTIFICATION IN THE MSP

Experiment 1 determined whether a listener could distinguish between two broadband noises having somewhat different spectra. Differences consisted of boosted or attenuated bands two-thirds of an octave in width. Two conditions were compared: a *fixed-location* condition, in which all of the sounds came from directly in front of the listener; and a *roved-location* condition, in which sounds randomly came from different locations in the MSP.

A. Environment

The listener was seated in an anechoic room (IAC 107840) with interior dimensions 3×4.3×2.4 m. There were five matched single-driver loudspeakers (Minimus 3.5) arrayed in the MSP.[1] The speakers were 122 cm from the listener's ears and equally spaced (45-deg increments) over the 180-deg span from directly in front to overhead, to directly behind. In order, they were called ''front,'' ''front–over,'' ''over,'' ''rear–over,'' and ''rear.''

B. Procedure

For each experimental run, there were two noises, called ''*A*'' and ''*B*,'' with different spectral structure. Each trial of a run consisted of a single presentation of a noise, either *A* or *B* selected randomly. The listener's task was to identify the noise as *A* or *B* and to indicate the choice by pressing a button on a response box. If the choice was correct, a lamp blinked on.

A run began with 15 training trials during which the listener learned to recognize the noises. On these trials, all stimuli came from the front speaker only. After training, the run continued with 50 data-collection trials. Data-collection trials were either fixed location or roved location. For fixed-location trials, the speaker in front was used exclusively; for roved-location trials, all five speakers were used with equal probability. Feedback was maintained throughout the data-collection trials.

Experimental runs were actually consecutive double runs, a fixed-location run followed by a roved-location run done with the same noise pair. The fixed-location run provided baseline data, as well as thorough training in the recognition of the noises prior to the rove. Each listener did three double runs for each $A-B$ noise pair. The listener was always aware of whether a run was fixed location or roved location.

C. Noise spectra

Stimuli for the experiment were Gaussian noises, 0.5-s long (10-ms rise/fall time), presented at a level of 72 dBA. The noises were passed through a computer-controlled 1/3-octave band equalizer that was set to create boosted or attenuated bands, 2/3-octave wide. A boosted band, referred to here as a ''bump,'' was 10 dB high; an attenuated band, referred to as a ''dip,'' was 10 dB deep.[2] Six noises with bumps and six noises with dips were created. Table I (column 1) gives their center frequencies. (Center frequencies were the same for the bumps and the dips.)

TABLE I. Center frequencies of 2/3-octave bumps and dips used in experiment 1 are given in the first column. A neighboring pair of bumps/dips, used as an $A-B$ pair in experiment 1, is characterized by a pair mean frequency, given in the second column.

Center frequency (kHz)	Pair mean frequency (kHz)
1.1	
	1.4
1.8	
	2.3
2.8	
	3.5
4.5	
	5.7
7.1	
	8.9
11.2	

The noises *A* and *B* presented on a run were either both bumps or both dips. *A* and *B* were always adjacent pairs from Table I, with their center frequencies separated by 2/3 of an octave. For example, noise *A* might be the 2/3-octave band centered on 4.5 kHz, and noise *B* the neighboring band centered on 7.1 kHz. Figure 1 shows how the equalizer was programed to create that pair of noises. A noise pair is referenced in this article by the geometric mean of its boosted (or attenuated) bands. The second column in Table I gives the mean frequency for each $A-B$ pair. In all, there were ten $A-B$ pairs (five with bumps and five with dips). The order of testing with these pairs was randomized differently for each listener.

D. Listeners

Four listeners were tested in this experiment and in the experiments that follow. Listener S1 was a male, age 19, with normal audiometric hearing in both ears and more than a year's experience in psychoacoustics experiments. Listener S2 was a female, age 21, with normal hearing in both ears and no prior experience. Listeners S3 and S4 were males, ages 45 and 59, respectively, with some bilateral high-

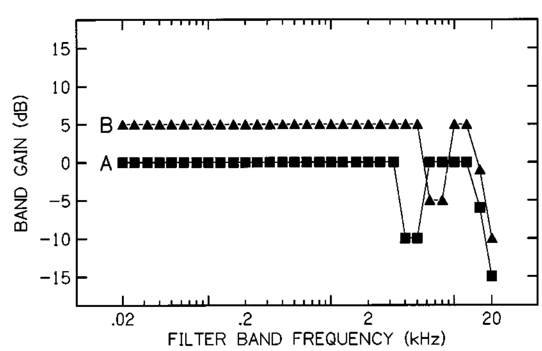

FIG. 1. A pair of noises used in experiment 1. Both noises have 2/3-octave dips as described in Table I. Noise *A* has a dip centered on 4.5 kHz, and noise *B* has a dip centered on 7.1 kHz. Noise *B* has been raised by 5 dB for clarity.

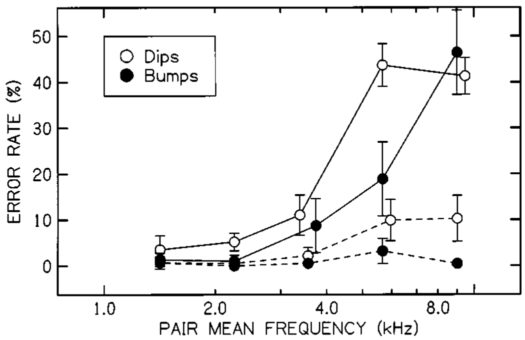

FIG. 2. Error rates for experiment 1. Roved-location runs are indicated by solid lines, fixed-location runs by dashed lines. The error bars show plus and minus one standard deviation for four listeners. (For the mean frequency near 9 kHz there were only three listeners.) Some data points have been shifted laterally for clarity.

frequency hearing loss. Both of these listeners had extensive listening experience. Listeners S1, S3, and S4 were the authors.

E. Results and discussion

The ten *A*–*B* pairs and four listeners led to 40 combinations both for fixed-location and for roved-location runs. There were two combinations where a listener failed to establish a reliable baseline for the fixed location (error rate greater than 25%). In those two cases, the roved-location runs were omitted.

The four listeners had very similar results overall. Figure 2 shows their mean identification error rate (±1 s.d.) for every noise pair, plotted as function of the pair mean frequency. Data are plotted separately for the bumps (filled symbols) and the dips (open symbols). Dashed lines connect results for fixed-location runs; solid lines connect results for roved-location runs.

The listeners generally found bumps easier to identify than dips [significantly lower overall error rate; $F(1,3) = 82.3$; $p < 0.001$], consistent with a previous report by Moore *et al.* (1989).

For bumps and dips, there was a significant interaction between the type of run—fixed location or roved location—and the mean frequency of a noise pair [$F(4,12) = 66.1$; $p < 0.001$]. When the spectral structure of the noises occurred at low frequency, near 1 or 2 kHz, identification was nearly perfect, no matter whether the location was fixed or roved. Listeners found the task easy to do. The spectral structure sometimes caused the noises to resemble vowel sounds, giving listeners an overlearned reference for identification. In other instances there was a pitch or timbre cue. As the frequency of the spectral structure increased, the error rate increased. Whereas the error rate for the fixed-location runs increased slightly, the error rate for the roved-location runs increased enormously, approaching the random guessing limit of 50% errors.

II. EXPERIMENT 2—ADDITIONAL SPECTRAL COMPARISONS

The frequency dependence found in experiment 1 was quite dramatic for all the listeners. Experiment 2 sought to determine whether a similar pattern would occur for a greater variety of stimuli. Otherwise, experiment 2 was the same as experiment 1.

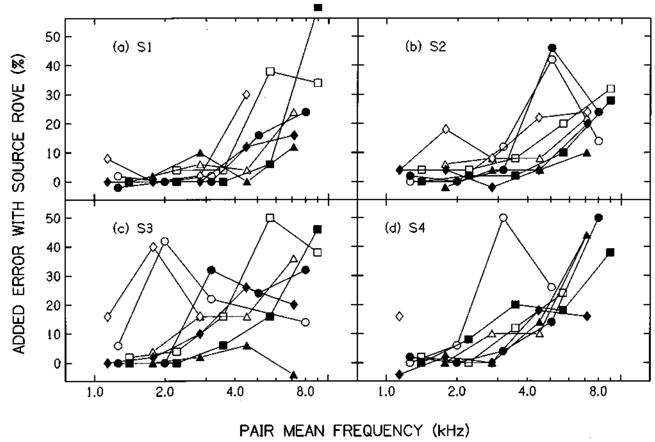

FIG. 3. Experiment 2 results. Added error is the increase in the percentage of errors in identification for pairs of noises when source location was roved compared to fixed-location presentation. Different panels show results for different listeners. Open symbols indicate pairs of spectral dips; filled symbols indicate pairs of spectral bumps. Different symbols represent different conditions, based on bandwidth and separation of the spectral structures as follows. Diamonds: one-third-octave bands separated by 2/3 octave; Circles: one-third-octave bands separated by 4/3 octave; Squares: two-thirds-octave bands separated by 2/3 octave; Triangles: two-thirds-octave bands separated by 4/3 octave.

A. Method

In addition to the 2/3-octave bumps and dips from the first experiment, experiment 2 included noises with bumps and dips that were only 1/3-octave wide. In addition to a 2/3-octave separation between the bumps (or dips) to be compared (as in experiment 1), experiment 2 included comparisons with 4/3-octave separation. Altogether, there were 38 *A*–*B* noise pairs in experiment 2 (19 with bumps, and 19 with dips). Because of the large number, experiment 2 was limited to one double run per *A*–*B* pair, a fixed-location run followed by a roved-location run.

B. Results

The results of experiment 2 are presented in Fig. 3. Each panel of the figure reports data for a different listener. In order to display the results for all *A*–*B* noise pairs at once, the figure has a different format here than in experiment 1. Figure 3 reports the *difference* in error rates between a roved-location run and the corresponding fixed-location run. Therefore, the data indicate the additional error percentage caused by roving the source location. Data points connected by lines are from the same stimulus type.[3]

Figure 3 shows that the effect of frequency on identification is quite general. For all the conditions and all the listeners, there was a strong tendency for the error rate to increase with increasing mean frequency of the spectral structure. The error rate increased steeply in the high frequencies. For every listener, the added error due to source rove was significantly greater for the noises with spectral structure above 4 kHz than it was for the noises with structure below 4 kHz [subject S1: $t(6)=4.19$; $p=0.006$; S2: $t(7)=5.12$; $p=0.001$; S3: $t(6)=2.60$; $p=0.041$; S4: $t(6)=7.21$; $p<0.001$].

C. Discussion

Experiment 2 verified the effect of frequency range on the ability of listeners to distinguish between noises with different spectral structures. When the relevant structure occurred at low frequency, 1–3 kHz, it did not matter whether the source position was fixed or randomized in the MSP; listeners could identify the noise. When the relevant spectral structure occurred above 4 kHz, listeners had difficulty distinguishing between noises when the source position was roved, and the difficulty continued to increase as the frequency range of the structure increased. With structure near 8 kHz, it became effectively impossible to identify some noises well enough to distinguish them when the position was roved.

Almost certainly these results relate to the frequency dependence of the head-related transfer function (HRTF). Prominent HRTF pinna cues in particular are known to appear at high frequencies (Shaw, 1966, 1982; Shaw and Teranishi, 1968). Pinna cues seem to have had an especially disruptive influence on listeners' ability to maintain a sense of source identity. Just why this occurred is unclear. The finding is, in fact, surprising from at least two points of view.

One of these views is motivated by Blauert's (1969–70) narrow-band noise experiments. Blauert found perceptual and physical evidence indicating that MSP directional bands are present all across the spectrum. Given this, it might be expected that roving the source location would have disrupting influences all across it as well. That is, one might expect to see poor performance on roved-location identification at low and mid frequencies, as well as at high frequencies.

Another view—the view that motivated the present study—is that listeners have some ability to factor out spectral variations that accompany changes in source location, that is, to perceptually re-equalize the spectrum according to its location before making any decision about source identity. On this view, the expected outcome of roved-location experiments is that identification performance should be good all across the spectrum, so long as the listener can localize the source. Based on the results of experiments 1 and 2, we tentatively conclude that listeners cannot in fact perform such a re-equalization. There is, however, another possible explanation for the results. It is possible that our listeners could not identify noises with high-frequency spectral structure because they could not localize them. If they did not know where the sounds were coming from, then they could not compensate for directionally dependent filtering. Experiment 3 was done to test for the possibility that listeners could not identify because they could not localize.

III. EXPERIMENT 3—IDENTIFICATION AND LOCALIZATION

The purpose of experiment 3 was to directly compare the ability to identify with the ability to localize. The identification part of experiment 3 was the same as experiment 1—three double runs with the source fixed in front or roved over five locations. After subjects completed the identification test, they went through a localization test using the same stimuli. The format of a localization test was identical to that of a roved-location identification test except that listeners had to answer a different question. Instead of deciding whether a sound was noise *A* or noise *B*, they had to decide where the sound came from. Localization decisions were reported by pressing one of five buttons on a control box. There was no feedback on localization runs.

The $A-B$ noise pairs tested here were selected to include two pairs that were easy for a listener to identify and three pairs that were hard. Easy and hard pairs were selected on the basis of a listener's performance in experiment 2. Specifically, they were the $A-B$ pairs that had previously produced the smallest (easy) and largest (hard) values of added error for each listener, per Fig. 3.[4]

TABLE II. Experiment 3: Percentage of correct identification (ID) and correct localization (LOC) responses for roved sources. For each listener, S1–S4, there were pairs of noises that were easy to identify and pairs that were hard to identify. The identification scores show that large difference. The difference in localization scores is much smaller. Chance performance for ID is 50%; chance performance for localization is 20%.

	Easy identification		Hard Identification	
Listener	ID	LOC	ID	LOC
S1	99	86	52	84
S2	96	95	59	87
S3	99	95	56	74
S4	99	60	60	54

A. Results

Table II summarizes the results of experiment 3. The mean percent-correct scores on the roved-location identification and localization tasks are reported for the easy (300 trials per listener) and hard (450 trials per listener) identification noises. Table II shows a dramatic difference in identification scores when the easy condition is compared with the hard condition. That is not surprising; the experiment parameters were chosen to give this result.[5] What is interesting is the comparison in localization performance for the easy and hard conditions. Listener S1 shows hardly any difference at all, less than two percentage points in localization accuracy. Listeners S2 and S4 show a modest difference in localization, less than ten percentage points, and listener S3 shows a large difference.

B. Discussion

1. The re-equalization hypothesis

The central question addressed by this experiment is whether listeners can use their knowledge about the location of a source to help identify a sound. When sources have random locations in the MSP, there are random spectral shape changes that make it difficult for listeners to identify sounds on the basis of their spectra (experiments 1 and 2). In principle, this difficulty should be eliminated if listeners can re-equalize the received spectra based on their knowledge of

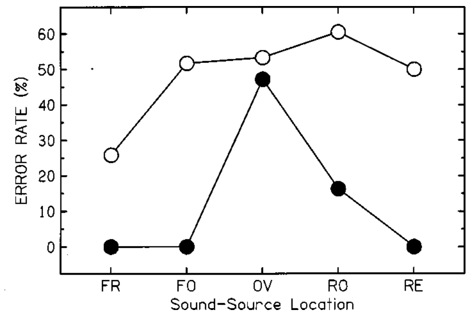

FIG. 4. Subject S1's error rates when localizing (filled symbols) and identifying (open symbols) the pair of 2/3-octave dips at 7.1 and 11.2 kHz. Errors rates are given for each of five sound-source locations in the MSP (FR=front, FO=front–over, OV=over, RO=rear–over, RE=rear).

source location. Such re-equalization could lead to a perceptual constancy whereby sounds can be identified regardless of their location in the MSP.

The results shown in Table II generally argue against the re-equalization hypothesis. The stimuli called ''hard'' are those for which the randomization of source location produced a large degradation in sound identification. Identification performance for those stimuli was poor indeed, quite near the 50 percent score expected for random guessing. However, the same stimuli were not poorly localized overall. To the contrary, Table II shows that localization performance was far better than the 20 percent score expected from random guessing and normally appreciably better than for identification. Therefore, the failure to identify cannot be attributed to a failure to localize.

Table II also reports a subject-specific difference that is relevant. Listener S4 was a poor sound localizer. His localization accuracy, although well above chance, was 30% below the mean accuracy for the others. Nevertheless, S4 identified sounds with an accuracy at or above the group mean for both the easy and hard identification runs. This result is additional evidence for a dissociation between localization and identification.

Still further evidence against the re-equalization hypothesis comes from a location-by-location comparison of identification and localization data. If re-equalization operates, then one might expect a listener to identify sounds accurately when they come from a location that is localized accurately and to identify sounds poorly when they come from a location that is mislocalized. In other words, identification performance should track localization performance, particularly for those locations where dramatic localization failure occurs. Three hard $A-B$ noise pairs for each of four listeners led to 12 opportunities to look for this. For each of those cases, we plotted the localization errors and identification errors at each location. Figure 4 shows an example.

Generally, localization errors showed sharp structure, with most of the errors occurring for only one or two sources, as in the example. For four cases, identification errors showed evidence of that same sharp structure. But for eight cases, including the example, identification errors were broadly distributed, occurring about equally often in each of the five loudspeaker locations, and so identification did not track localization. Therefore, most of the data argue against the concept of re-equalization.

TABLE III. A tally of the number of sound localization errors of different sizes that each subject made in experiment 3.

Listener	Localization error size			
	45 deg	90 deg	135 deg	180 deg
S1	108	0	2	1
S2	73	2	0	0
S3	130	3	0	0
S4	215	76	20	17

In sum, experiment 3 provided three kinds of evidence that sound identification and sound localization are largely dissociated, and hence three kinds of evidence against the re-equalization hypothesis. (i) There was evidence of dissociation across stimuli. For most listeners, localization accuracy was about the same for sounds that were easy to identify and sounds that were hard to identify. (ii) There was evidence of dissociation across listeners. Identification performance was more or less the same for all four listeners, despite the fact that one of them was decidedly poor at MSP sound localization, compared to the other three. (iii) There was evidence of dissociation across locations. Identification errors occurred about equally often for all five source locations, but localization errors occurred focally at one or two locations only.

2. Differences in sound localization ability

Table II shows some clear differences among the four subjects regarding their sound-localization abilities. Most especially, subject S4 was distinctly poorer than the others at sound localization. We wondered whether these subject differences were specific to the localization of bump/dip noises of the sort generated for the experiment, or more general. They proved to be more general. In a follow-up experiment, we presented white-noise stimuli randomly from the five loudspeaker locations and had the subjects make localization judgments, as in the main experiment. Listeners S1, S2, S3, and S4 correctly localized the white noise on 84%, 97%, 94%, and 51% of the trials, respectively. These results track the subject dependence seen in Table II and show that the differences among subjects in the localization part of experiment 3 are unrelated to the special spectral structures.

3. Localization error size

Table III shows an analysis of the localization errors made in experiment 3. Each subject's errors are broken out according to their size (in deg). The subjects differed substantially in terms of the total number of errors that they made, but they were in general agreement about error size. When errors occurred they were almost always 45-deg or 90-deg errors.

It is notable that front-to-back errors (180 deg) were few. For listeners S1, S2, S3, and S4, the numbers of front-to-back errors were: 1 in 289 (front or rear presentation) trials, 0 in 296 trials, 0 in 305 trials, and 17 in 298 trials, respectively. Thus, for listeners with normal localization

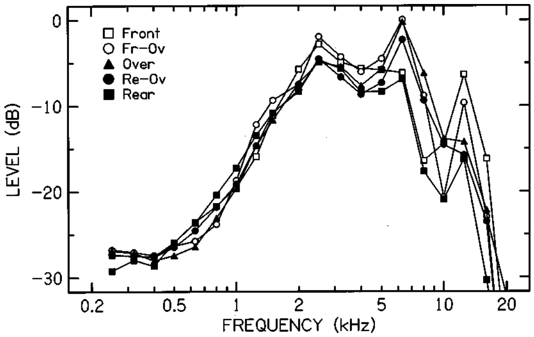

FIG. 5. Spectral density for a white noise presented from each of the five sources used in the identification and localization experiments. Measurements were made with the left ear of a KEMAR and are given in one-third-octave bands.

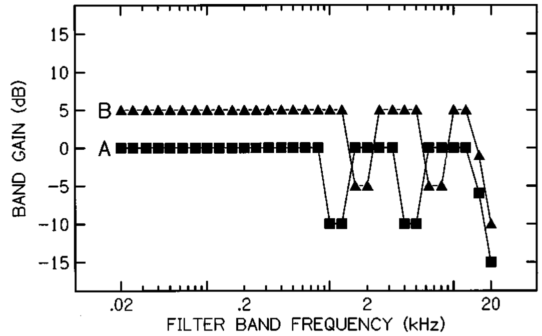

FIG. 6. A pair of noises used in experiment 4 with noise *B* raised by 5 dB for clarity. Each noise includes a high-frequency dip and a low-frequency dip.

abilities (S1–S3) there was only one front-to-back confusion in 890 trials with either the front or the rear speaker as the source. What seems remarkable about this result is that transfer functions measured in the ear canal or with a dummy head typically show that HRTFs are similar for sources at front and back locations. Therefore, front-to-back confusions are expected. The fact that our listeners made so few front-to-back errors seems paradoxical. Evidently, when presented with real-world sources, our listeners were able to make the fine distinctions necessary to avoid large localization errors. Further consideration of subjects' localization errors, and their relationship to HRTF features, is presented in the Appendix to this article.

4. Re-equalization again

Experiment 3 showed a number of instances in which listeners identified sounds poorly, but localized them accurately. We took these to be evidence against the re-equalization hypothesis. A counter argument would be that the spectral changes that mediate localization might be larger and more salient than the spectral differences between the sounds that listeners had to identify in our experiments. Therefore, an ability to localize might not necessarily imply an ability to identify. To get quantitative information on this point, we made recordings with a Knowles Electronics Manikin for Acoustic Research (KEMAR) placed at the listener's position in the anechoic room. Binaural recordings were made for white noise sent to each of the five loudspeakers.[6]

The left- and right-ear recordings were not ideal HRTFs; they included the response of the loud-speakers and stimulus roll-off above 12.5 kHz associated with the equalizer program (Fig. 1). The advantage of the recordings was that they were made in the conditions experienced by our listeners. Responses measured for the manikin's left ear are shown in Fig. 5. The right-ear response functions were similar, though not identical. (Left and right KEMAR ears are not exact mirror images.) The response functions for both ears showed a number of features that are expected for HRTFs. For example, there was a broad outer-ear resonance above 2 kHz. Sources overhead lead to a peak near 6 or 7 kHz. The response to a source in the rear was greater than the response to a source in front over a broad region around 1 kHz.

Figure 5 shows that the re-equalization counter argument is incorrect. Spectral differences between sources 45 deg apart (or, better, 90 deg apart) are not greater than the 10-dB bumps and dips that we introduced to try to make sounds distinguishable from one another. The fact that listeners can localize to within 90 deg with nearly 100-percent reliability means that they are sensitive to rather subtle spectral differences, smaller than the stimulus distortions of 10 dB in 1/3- or 2/3-octave bands. Therefore, there is no quantitative physical argument against the conclusion that listeners are unable to re-equalize.

IV. EXPERIMENT 4—COMPLEX STRUCTURE

Perceptual experiments 1 through 3 above asked listeners to identify noises characterized by spectral structure in a single frequency region and found evidence of substantial confusion when the structure was at high frequency and the source location was roved. Most ecologically significant sounds do not have their spectral structure confined to a single region. For example, speech has formant bands all across the spectrum. To learn more about sounds like speech, experiment 4 had listeners identify noises with complex structure.

A. Method

The experiment was done according to the protocol of experiment 1, with three double runs for each A–B noise pair. Complex noises of two types were presented. There were *double-dip* noises that combined a 2/3-octave dip at low frequency with a 2/3-octave dip at high frequency, and there were *double-bump* noises that combined low- and high-frequency bumps. For each listener, we found a pair of low-frequency dips/bumps that was well identified in experiment 1, and a pair of high-frequency dips/bumps that was poorly identified. We then combined them in low–high pairs to make the complex noise stimuli for this experiment.[7]

Figure 6 shows an example. Noise A is a double-dip noise that marries dips at 1.1 and 4.5 kHz. Noise B is a double-dip noise that marries dips at 1.8 and 7.1 kHz. We always combined the lower-frequency member of a low-frequency pair with the lower-frequency member of a high-frequency pair to derive one double-dip/bump stimulus, and then combined the higher-frequency members of each pair to

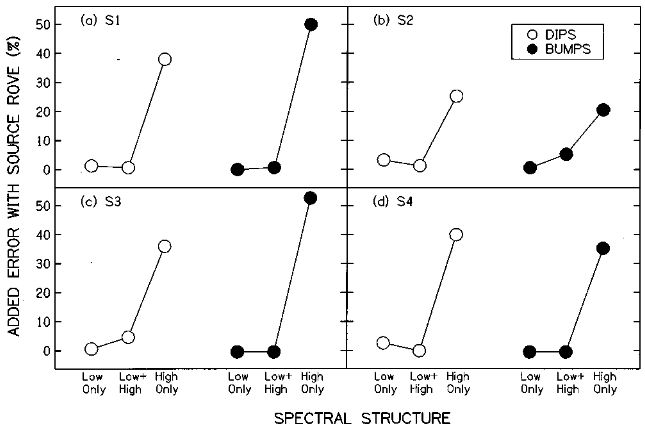

FIG. 7. Experiment 4 results for listeners S1–S4 in panels (a)–(d), respectively. As in Fig. 3, the ordinate shows added error when a source is roved.

derive the other stimulus, as in the example. Each listener heard one double-dip pair and one double-bump pair in the experiment.

B. Results and discussion

The results of experiment 4 are shown in Fig. 7. The added error due to source rove is plotted there for three kinds of spectral structure. There is a *low-only* data point from experiment 1 with an $A-B$ noise pair consisting of low-frequency bumps/dips. There is a *high-only* data point, also from experiment 1, with an $A-B$ noise pair consisting of high-frequency bumps/dips. Finally, there is a *low+high* data point from the present experiment using an $A-B$ pair of complex noises that combined the low-only and high-only spectral features.

For every listener (different figure panels) and for both bumps (filled symbols) and dips (open symbols), the outcome was the same. Roving location caused little or no added error for identification of complex noises with low +high spectral structure. They were identified as accurately under rove as were the low-only noises, and dramatically better than the high-only noises. We conclude that when source location varies, listeners can maintain constancy in their identification of complex sounds by latching onto low-frequency structure in the sound spectrum. Their grip on low-frequency structure appears to be firm, firm enough to anchor perception of what would be otherwise ambiguous high-frequency structural components.

V. CONCLUSIONS

Experiments 1 and 2 tested listeners' ability to identify broadband sounds with different spectral structures. The experiments showed that when the source location of the noises was fixed in space, listeners could usually identify the noises successfully. Listeners were somewhat more successful at identifying noises with low-frequency structure (1 to 3 kHz) than noises with high-frequency structure (above 4 kHz), possibly because low-frequency structure often led to vowel-like sounds. However, the difference between low and high frequencies was not great. By contrast, when source location was randomized among five locations in the median sagittal plane (MSP), listeners were still able to identify noises with low-frequency spectral structure, but their ability to identify high-frequency structure fell almost to the chance level.

Experiment 3 compared the listeners' ability to identify noises with their ability to localize them. It showed that the difference in identification between noises with low- and high-frequency structure cannot be explained by any corresponding differences in MSP sound localization. Three listeners localized all of the noises with a high accuracy, no matter what the spectral structure. A fourth listener localized less well than the others, but substantially better than chance, and at about the same level no matter whether the noises to be localized had low-frequency or high-frequency structure.

Overall, the pattern of localization and identification results for experiment 3 argued against a *re-equalization* hypothesis, which says that listeners can compensate for spectral distortions due to their HRTFs before deciding about the identity of incoming sounds. Because they can localize sounds on the basis of their spectral signatures, it must be presumed that listeners are aware, at some level, of the directionally dependent filtering that takes place. The information provided by a known source location could, in principle, be used to re-equalize the spectra to enable identification. Our results nevertheless showed that listeners cannot do that, even after substantial practice. The failure to distinguish among high-frequency spectral structures with roving location was dramatic and quite similar for all our listeners in experiments 1–3. The failure was also persistent. It persisted throughout a run, showing up equally at its beginning and end, and when there were multiple identification runs with a particular noise pair it persisted across the runs.

Macpherson (1995) also concluded that listeners cannot compensate for their location-dependent HRTFs. In a headphone profile analysis experiment, Macpherson caused a listener's identification task to be difficult by filtering the signals with assorted HRTFs. He found that identification performance was not best when the HRTFs were intended to provide a realistic representation of source location. Rather, it was best for certain mismatched left–right pairs of HRTFs that did not correspond to any location. Although his experiments did not measure perceived localization directly, the results are consistent with the dissociation between identification and localization found here.

Finally, experiment 4 showed that listeners can distinguish between complex noises that include spectral structure at both high and low frequencies. This is true whether the source location is fixed or roved in the MSP. The significance of this result is that although high-frequency structure may be perturbed to the point of confusion by the roving of location, listeners can ignore that confusion and identify complex sounds on the basis of the low-frequency structure alone.

ACKNOWLEDGMENTS

This project was supported by the National Institutes of Health through NIDCD Grant No. DC00181. Funds were also provided by the National Science Foundation Research Experience for Undergraduates (REU) Program in the Michigan State University Department of Physics and Astronomy.

We are grateful for the help of Melanie Koskamp, a participant in the REU program, and to Ewan Macpherson for useful discussions.

APPENDIX: MSP SOUND LOCALIZATION AND HRTFS

In experiment 3, three listeners (S1–S3) accurately localized noises with a variety of bump and dip structures most of the time, and when they did make mistakes their errors were almost always confusions with a nearest-neighbor location in the loudspeaker array (i.e., 45-deg errors; see Table III). A fourth listener (S4) made more localization errors than the others, but those errors too were predominantly confusions with a nearest neighbor or, to a lesser extent, a next-nearest neighbor.

A notable feature of the listeners' errors was that they were not at all uniformly distributed among the five sound-source locations. For listeners S1–S3, adding particular bumps or dips to the spectrum often caused the overhead location, or the rear–over location, or sometimes both, to be incorrectly localized. The failures for these one or two locations were dramatic and systematic. For example, putting a 10-dB dip in the two-thirds-octave band at 7.1 kHz caused listener S1 to miss 7 out of 14 overhead localizations, and every missed overhead was mislocalized as rear–over. The identical error was made by listener S2 on 6 out of 10 trials for this stimulus. For listener S3, this stimulus caused rear–over to be mislocalized as rear on 9 out of the 11 trials.

It is interesting to ask whether these highly systematic localization errors can be understood on the basis of deformations in the apparent HRTFs. Accordingly, we compared the errors with the KEMAR HRTF measurements that were made for experiment 3 (Fig. 5). The comparison is not ideal because KEMAR's HRTFs and those of an individual listener can differ, sometimes substantially. We nevertheless found several notable points of correspondence with the localization results. The failure of listeners S1 and S2 to correctly localize the overhead source in the presence of a 10-dB dip at 7.1 kHz seemed easy to understand. Figure 5 shows that the overhead position is most favored at 8 kHz. It also shows what Butler and Musicant (1993) have called a ''covert peak,'' a peak at the overhead position in a plot of intensity vs position for a given frequency of 8 kHz. Putting a dip in the 7.1-kHz band makes the overhead source look more like the other sources. It is less evident why this spectral change should cause the image of the overhead source always to appear at the rear–over position. This is probably caused by similarities between rear and rear–overhead at lower-frequency regions of the spectrum. For listener S3, the rear–over source reliably appeared at the rear position, and this result is also consistent with the change in the tall peak of an apparent HRTF.

Given the above results with a 10-dB spectral dip at 7.1 kHz, it might be predicted that its opposite, namely a 10-dB bump in that band, would lead to errors of the opposite kind. As it turned out, both listeners S1 and S2 had that bump on their menu, and both experienced the erroneous localizations as expected. For listener S1, the rear–over location was missed on 8 out of 15 trials and the incorrect localization always went to overhead. For listener S3, the rear location was missed on 10 of 15 trials and the incorrect localization always went to rear–over. Thus, for both S1 and S3, the spectral dip and the complementary spectral bump produced errors of exactly opposite kinds.

Other systematic errors are more difficult to explain in terms of the KEMAR data. Adding a spectral bump in the two-thirds-octave band at 11.2 kHz caused listener S1 to hear the overhead source at the rear–over position on 10 out of 14 trials. HRTFs measured with the KEMAR do not show any kind of peak for the rear–over location in this frequency region. One might understand the judgments of S1 by assuming that he uses the high-frequency region above 10 kHz as a reference against which other peaks may be judged in the manner of profile analysis (Green, 1988). Boosting this high-frequency region would make a peak near 8 kHz due to an overhead source seem relatively less prominent and possibly cause the overhead source to be incorrectly localized.

Although it is possible to explain certain systematic errors in localization in terms of HRTFs and the spectral distortions deliberately introduced in our experiments, an arguably larger issue is to explain why our listeners localized as well as they did. Despite large spectral distortions in frequency ranges known to be important for localization in the MSP, their localization performance was good overall. It seems likely that one key to understanding this behavior is that our spectral distortions occurred in isolated frequency regions. Our results suggest that listeners do not rely entirely on isolated spectral features, such as the frequencies of high-frequency peaks or valleys. Instead, they may use those features in combination with spectral details from a wide range of frequencies. Alternatively, it may be that localization depends on a rather coarse analysis of the HRTF, as proposed in several recent models (Fuzessery, 1986; Zakarauskas and Cynader, 1993; Macpherson, 1997). If this is the case, then local distortions, like the ones introduced here, could leave much of the spectral information intact.

[1]The five loudspeakers were selected from a batch of 79 on the basis of their best-matched frequency response, as determined by an automated comparison made at 76 frequencies from 125 to 11 000 Hz.

[2]The equalizer was a model IEQ made by Applied Research and Technology. This equalizer has a ''smart curve'' feature that makes the spectral shape of a bump or dip close to the rectangular ideal by automatically adjusting the gains of adjacent one-third-octave bands.

[3]The caption of Fig. 3 gives the different combinations of bump/dip bandwidths and separations between spectral features that were included in experiment 2. Error rates on the fixed-location runs provided a baseline for the ''added error due to source rove'' computation. The fixed-location rates for experiment 2 were similar to those observed in experiment 1 (see Fig. 1). For bumps, fixed-location error rates were generally less than 5% at all frequencies. For dips, rates were somewhat higher, 5%–10% at frequencies below 4 kHz, and 5%–15% above 4 kHz. As in experiment 1, if a listener had an error rate greater than 25% for a fixed-location run, the corresponding roved-location run was omitted. Only spectral dips led to such failures in experiment 2; bumps never did. Of 40 combinations of listener and one-third-octave dips, there were 11 failures. Of 36 combinations of listener and two-thirds-octave dips, there was one failure.

[4]The ''easy'' and ''hard'' $A-B$ pairs selected for each subject were as follows. S1-easy: 2/3-octave dips at 1.1 and 1.8 kHz, 2/3-octave bumps at 1.8 and 2.8 kHz. S1-hard: 2/3-octave bumps at 7.1 and 11.2 kHz, 2/3-octave dips at 4.5 and 7.1 kHz, 2/3-octave dips at 7.1 and 11.2 kHz. S2-easy: 2/3-octave dips at 1.8 and 2.8 kHz, 2/3-octave bumps at 1.1 and 1.8 kHz. S2-hard: 1/3-octave bumps at 4.0 and 6.3 kHz, 1/3-octave dips at 4.0 and 6.3 kHz, 2/3-octave dips at 7.1 and 11.2 kHz. S3-easy: 2/3-octave dips

at 1.1 and 1.8 kHz, 2/3-octave bumps at 1.8 and 2.8 kHz. S3-hard: 2/3-octave dips at 7.1 and 11.2 kHz, 2/3-octave dips at 4.5 and 7.1 kHz, 2/3-octave bumps at 7.1 and 11.2 kHz. S4-easy: 2/3-octave dips at 1.1 and 1.8 kHz, 2/3-octave bumps at 1.1 and 1.8 kHz. S4-hard: 1/3-octave bumps at 6.3 and 10.0 kHz, 1/3-octave dips at 2.5 and 4.0 kHz, 2/3-octave dips at 4.5 and 11.2 kHz.

[5]The difference between the easy identification and hard identification runs was equally dramatic when measured as added error due to source rove. The average added error was 1.1% for the easy runs and 36.1% for the hard ones.

[6]The KEMAR (Knowles Electronics Manikin for Acoustics Research) was fitted with two large ears, Knowles DB-065 and DB-066, and Zwislocki couplers. The couplers included Etymotic ER-11 half-inch microphones with accompanying preamplifiers set to a ''flat'' response. Output signals from KEMAR's left and right ears were analyzed in 30 one-third-octave bands by a Stanford Research Systems SR-760 FFT spectrum analyzer.

[7]S1 double-dips: 1.1 and 7.1 kHz, 1.8 and 11.2 kHz. S1 double-bumps: 1.8 and 7.1 kHz, 2.8 and 11.2 kHz. S2 double-dips: 1.8 and 7.1 kHz, 2.8 and 11.2 kHz. S2 double-bumps: 1.1 and 4.5 kHz, 1.8 and 7.1 kHz. S3 double-dips: 1.1 and 4.5 kHz, 1.8 and 7.1 kHz. S3 double-bumps: 1.8 and 7.1 kHz, 2.8 and 11.2 kHz. S4 double-dips: 1.1 and 4.5 kHz, 1.8 and 7.1 kHz. S4 double-bumps: 1.1 and 7.1 kHz, 1.8 and 11.2 kHz.

Blauert, J. (**1969–70**). ''Sound localization in the median plane,'' Acustica **22**, 205–213.

Butler, R. A., and Belundiuk, K. (**1977**). ''Spectral cues utilized in the location of sound in the median sagittal plane,'' J. Acoust. Soc. Am. **61**, 1264–1269.

Butler, R. A., and Musicant, A. M. (**1993**). ''Binaural localization: Influence of stimulus frequency and the linkage to covert peak areas,'' Hearing Res. **67**, 220–229.

Durlach, N. I., and Colburn, H. S. (**1978**). ''Binaural Phenomena,'' in *Handbook of Perception*, Vol. IV, edited by E. Carterette (Academic, New York).

Fuzessery, Z. M. (**1986**). ''Speculations on the role of frequency in sound localization,'' Brain Behav. Evol. **28**, 95–108.

Green, D. M. (**1988**). *Profile Analysis Auditory-Intensity Discrimination*, Oxford Psychology Series No. 13 (Oxford, New York).

Hebrank, J., and Wright, D. (**1974**). ''Spectral cues used in the localization of sound sources on the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834.

Macpherson, E. A. (**1995**). ''Source spectrum recovery at different spatial locations,'' J. Acoust. Soc. Am. **98**, 2946(A).

Macpherson, E. A. (**1997**). ''A comparison of spectral correlation and local feature-matching models of pinna cue processing,'' J. Acoust. Soc. Am. **101**, 3104(A).

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.

Moore, B. C., Oldfield, S. R., and Dooley, G. J. (**1989**). ''Detection and discrimination of spectral peaks and notches at 1 and 8 kHz,'' J. Acoust. Soc. Am. **85**, 820–836.

Plenge, G., and Brunschen, G. (**1971**). ''*A priori* knowledge of the signal when determining the direction of speech in the median plane,'' Proceedings of the Seventh International Congress on Acoustics, 19, H10.

Shaw, E. A. G. (**1966**). ''Ear canal pressure generated by a free sound field,'' J. Acoust. Soc. Am. **39**, 465–470.

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).

Shaw, E. A. G., and Teranishi, R. (**1968**). ''Sound pressure generated in an external ear replica and real human ears by a nearby point source,'' J. Acoust. Soc. Am. **44**, 240–249.

Wightman, F. L., and Kistler, D. J. (**1989a**). ''Headphone simulation of free field listening I,'' J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). ''Headphone simulation of free field listening II,'' J. Acoust. Soc. Am. **85**, 868–878.

Zakarauskas, P., and Cynader, M. S. (**1993**). ''A computational theory of spectral cue localization,'' J. Acoust. Soc. Am. **94**, 1323–1331.