F. Briolle

Localization in an Echoic Environment

E.R. Hafter, K. Saberi, E.R. Jensen and F. Briolle

Department of Psychology, University of California, Berkeley, CA 94720, USA

ABSTRACT

The effects of echoes on azimuthal discrimination and localization in the lateral plane were studied using single wide band clicks. Over much of the frontal quadrant, echoes decreased discrimination by no more than a factor of 2, though effects were greater near the aural axis. In general, localizations were of perceptually fused images closer to the source than the echo, but for large echo delays, the image sometimes broke into parts.

KEYWORDS

Localization; Multi-directional field; Echoes; Spatial Discrimination; Precedence

PRECEDENCE: HIGHER ORDER PROCESSING

The term binaural "precedence" describes the observation that a sound seems to emanate from the direction of its source, despite the presence of echoes. It has been used to support the argument that echoes are suppressed. Often cited in this regard is a report by Wallach *et al.* (1948), who presented two dichotic clicks, each with a different Interaural Difference of Time (IDT), and found that lateralization depended much more on the first click than the second. More complicated but nevertheless similar results were found by Gaskell (1983) with Interaural Differences of Level (IDL). Excellent reviews of precedence are found in Blauert (1983) and Zurek (1987).

Perrott *et al.* (1989) offer a direct test of the echo-suppression hypothesis. They presented signals from a loudspeaker on the midline followed by simulated echoes placed either to the left or right. Even with sounds fused into a single image, listeners were able to report the side of the pseudo echoes, showing that all lateral information had not been lost. Saberi and Perrott (1990) extended this to headphones, showing that highly trained listeners could recognize the side of a pseudo echo.

A more cognitive view of precedence that sees echoes as being ignored rather than suppressed is discussed in Hafter *et al.* (1988a) in a review of "binaural adaptation." The adaptation paradigm asks listeners to detect IDTs or IDLs in the envelope of a complex signal and analyzes the amount of interaural information transferred by each peak in the envelope. What is found is that the effectiveness of interaural cues declines beyond the stimulus onset by an amount inversely related to the modulation frequency. At first, this would seem to be an ideal candidate for the mechanism of precedence, but Hafter *et al.* (1988a) argue that adaptation is a low-level process of the peripheral auditory system, whereas precedence requires top-down processing. Their contention stems from studies of the recovery from adaptation which show that virtually any spectral change occurring during the adapted state is sufficient to restart binaural processing, inducing the system to resample its spatial environment (Hafter and Buell, 1990). What, then, of sounds whose durations are such that they overlap with their echoes? According to the restart phenomenon, a spectral change detected during the overlap should lead to resampling of the interaural cues, now the sum of the first waveform *plus* its echo, producing a new apparent azimuth based on the combination. Thus, one might expect precedence to fail in fields with multiple sources such as the cocktail party environment. Yet, it is the general consensus, albeit without conclusive evidence, that precedence does work in such situations, and it is this ineffectiveness of overlapping echoes that led Hafter *et al.* (1988a) to deduce that precedence results from a central process that sorts out the inputs, deciding which are sources and which are echoes. Related to this, they note that highly contextual stimuli such as speech and music produce precedence across much longer echo delays (Wallach *et al.*, 1949).

A demonstration by Clifton (1987) lends further support to the cognitive view of precedence. She presented a stream of clicks from two loudspeakers, one on either side of the midline. When clicks from the left led those from the right by a few ms, the stimuli fused to produce the perception of a single stream on the left; this is as predicted by all models of precedence. When the order of occurrence was suddenly reversed so that clicks on the right now led, there was a transitional period of a few pairs after which the percept was of a fused stream, on the right. The unusual result occured during the transitional period, when listeners reported hearing two distinct streams, one on each side. It is what one might expect from a cognitive process that, having committed to a solution in which clicks on the right are echoes of those on the left, cannot deal with a paradox in which *echoes lead*. Thus, for a few clicks, the system perseverates, believing the left clicks still to be sources, and this leaves no choice other than segregation into separate streams. After a bit, the processor realizes that the left clicks are now legitimate echoes of the right ones and the transition ends. We believe that this is a good example of what Hartmann and Rakerd (1989) have called the "plausibility hypothesis" in discussion of another onset phenomenon, the Franssen effect. The hypothesis basically says that binaural processing of an ambigious stimulus relies on computation and acceptance of the most plausible scenario.

LOCALIZATION IN THE MULTI-DIRECTIONAL FIELD

The long-term goal of our research is an understanding of the effects of echoes on many features of spatial hearing, but the present paper concentrates on two basic abilities: discrimination between sources in the horizontal dimension and localization. The signals are single, wide band clicks. One might ask, "Why not just simulate multi-directional listening through earphones using head-related transfer functions (HRTFs) to produce a 'virtual environment' (Poesselt *et al.* 1986; Wightman and Kistler, 1989)?" In reply, we note that it is a common observation that such simulations do not sound the same as sounds heard in the real multi-directional field; rather, they tend to appear closer to and generally in back of the head. It is true that the sense of externalization can be improved by modifying the HRTFs in response to movements of the head (Wenzel, 1991), and one suspects that increasingly sophisticated virtual environments will produce perceptions approaching those experienced in nature. Nevertheless, as these new techniques are developed, testing them will require comparisons to a database of normative results actually obtained in open-air listening. Thus, we feel that if we are to fully understand the role of echoes in realistic environments, we must first measure their effects in the multi-directional field. We should point out that use of complex stimuli in the multi-directional field raises problems of its own, and some of those will be discussed here.

GENERAL PROCEDURE

The Listening Environment: Experiments were conducted in an 8.3 x 5.4 x 4.0 m anechoic chamber with the listener placed along the long axis with his head 2.18 m from the back wall. An array of up to 8 speakers could be placed at azimuths ranging up to $\pm 40°$ on either side of the midline. For tests reported here, speakers were at the height of the subject's ears atop sand-filled stalks which were themselves mounted on rails. After placing a speaker at the desired location, its directional spectral effects were minimized by pointing the speaker toward the subject with a laser. Sounds were always presented at least 2.75 m from the subject to ensure stimulation from the acoustic far-field. Rails and speakers were hidden from view behind acoustically transparent white curtains.

Head placement: Study of localization requires accurate placement of the head and bite-bars raise questions about the possible involvement of middle-ear muscles. Earlier we used an optical device for fixing the listener's head during testing, but here, a magnetic-field device (Polhemus Isotrack

System) read the location and rotation of a sensor atop the listener's head and conveyed errors to the subject through an optical display. Stimulus presentation required that displacement errors be corrected.

Speakers and digital filters: Because of the great accuracy of localization, comparisons between sources requires relatively small loudspeakers; we have settled on Audax MHD12P25 FSM-SQ, 9-cm, acoustic-suspension mid-range speakers. The use of wide band stimuli requires that variations both within and between speakers be eliminated, and for this, the spectral response of each loudspeaker was flattened in both amplitude and phase by complex digital filters (Hafter *et al.*, 1988b). For generation of these filters, each speaker was placed 40.6 cm from a B & K microphone and the speaker is sounded with a computer-generated impulse. The cross-spectrum between the impulse and the speaker's response as computed in a spectrum analyzer was used to create a 512-point complex filter. In deference to reflective limitations set by the anechoic chamber, signals were kept within the range from .8 to 8 kHz, thus reducing demands on the dynamic ranges of the speakers and filters. Appropriately filtered impulse responses showed maximum variation across the entire frequency range of about 0.2 dB, with comparably small variation in phase.

Speaker variation: A technical problem arose when we found that speaker-correction filters could not be relied upon for long periods. Preparing for a morning experiment, we found that flat spectral responses of the night before now varied by as much as 2 dB. For a good profile listener (Green, 1988), this could allow easy discrimination between speakers based on cues other than azimuth. The problem seemed to arise from changes of temperature and humidity in the anechoic chamber. In subsequent testing, we have found that filters created at the beginning of a 4-hour session are safe, and have adopted that precaution. One cannot say whether our experience is universal, but we caution those working in the open field to be aware of speaker variation.

Stimuli: The stimuli used here were clicks made by passing an impulse through a Gaussian filter with a center frequency of 4 kHz and 1-sigma points of ± 700 Hz. The level of a click without echo was approximately 63 dB SL. Note that while these clicks were reasonably wide in bandwidth, they contained no energy in the low-frequency region normally thought to be localized on the basis of IDT. Each signal consisted of a source, usually followed by a single echo. The midline *source-speaker* was physically located 4.82 m from the listener. For other sources along the straight rails, the computer corrected the levels to simulate a distance of 4.82 m. Simulated echoic surfaces could be placed at arbitrarily chosen distances and azimuths. This was done by placing *echo-speakers* along the chosen azimuth and calculating the appropriate amplitudes and delays by assuming specular reflections with coefficients of 100%.

EXPERIMENT 1: DISCRIMINATION

Here we examined Minimum Audible Angles (MAAs) in a two-alternative, forced-choice (2AFC) task. Each trial employed three speakers, two for the sources and one for the simulated echoes. The subject's chair was rotated to allow for placement of the source speakers at different basal azimuths relative to the sagittal midline. No feedback was given.

Roving Echoes: In addition to individual differences between speakers, another potential confound in 2AFC discrimination is the fact that both the amplitude and delay of the echo depend on the location of the sources. We addressed this problem by randomly roving the site of the echoic surface from trial to trial. A preliminary study (Hafter and Briolle, 1989) found that roving echoes did not adversely affect azimuthal sensitivity. Indeed, untrained subjects often performed better with roving echoes than with fixed, perhaps because roving helped to separate "figure" from "ground." The sixteen potential echoic surfaces represented 4 azimuths and 4 distances.

RESULTS

As in reverberant rooms (Rakard and Hartmann, 1985), single echoes here reduced the discriminability of location. Fig. 1 shows MAAs for a single subject, both with and without echoes, as functions of the basal azimuth of the sources. Clearly, echoes decreased discriminability, implying an increase in the lateral extent of a source-plus-echo, what Blauert (1983) has called the "localization blur." Also, similar to the familiar pattern found with tones (Mills, 1958), the MAAs increased as the source moved away from the midline, as did the effects of echoes.

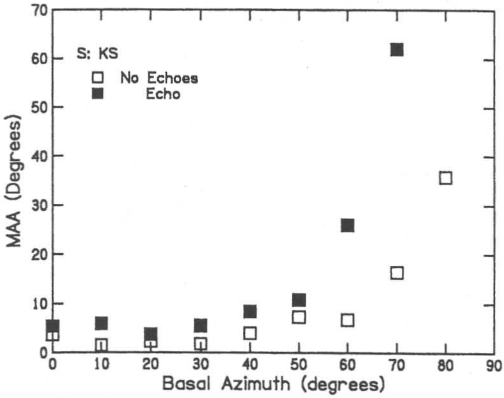

Fig. 1. Minimum Audible Angles from a single subject obtained both with and without echoes. Basal azimuths are relative to the saggital midline.

EXPERIMENT 2: LOCALIZATION

While MAAs offer a measure of the accuracy of localization, they cannot tell the whole story because they do not tell how echoes affect the perceived location. For this reason, subjects were asked to point to the source of the signal. A set of 30 trials randomly presented the source from one of 6 locations, each time paired with one of 5 echoes. As shown in Fig. 2, sources and echoes shared azimuths. Simulated distances from source to subject, D(s), were 2.54 m and, within each session, simulated distances from the echoic surfaces to the subject were D(e).

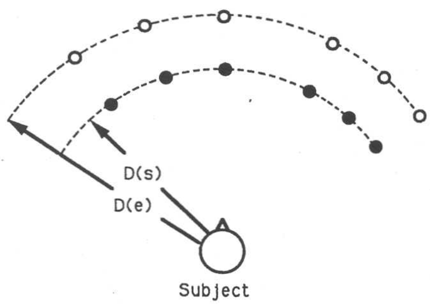

Fig. 2. Stimulus configuration for Expt. 2. There were six potential source locations (black) at simulated distances of D(s) and five potential echoic reflective surfaces (white) at simulated distances of D(e).

Pointers: Middlebrooks *et al.* (1989) used a head tracker to measure the azimuth of single sources by having the subject look toward the image, but we were concerned that source-plus-echo localizations might be too fragile to withstand head movements made during the response. Thus, subjects here pointed to the location of a fused image or, if there were multiple images, separately to each with a laser mounted on orthogonal potentiometers monitored by the computer. Signals came 1/s and listeners responded whenever they wished. Measures were also made without echoes.

RESULTS

Fig. 3 shows localizations of three subjects obtained *without* echoes. Note that while apparent locations were close to the correct azimuths, they show consistent biases. Blauert (1983) and Butler *et al.* (1990) have discussed biases imposed by monaural pinna-cues, especially on localization of narrow band signals and in the vertical dimension. We were surprised to find such biases with reasonably wide band stimuli; however, the differences seen between three individuals suggest that the biases were indeed due to HRTFs. Wightman (1991) suggests that these biases might disappear if the clicks contained low frequencies; future measures will test his conjecture.

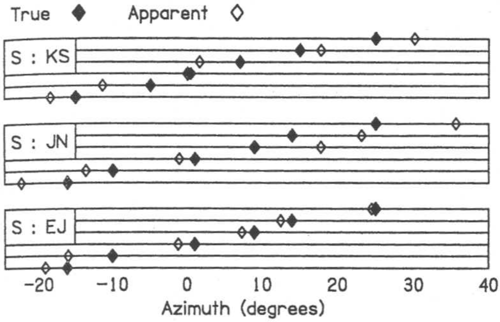

Fig. 3. True and apparent locations obtained without echoes. The basal azimuth was 0°.

Localization data from conditions with echoes are shown in Fig. 4, corrected to offset the biases described above. Each line depicts results with a single source location (shown by solid triangle) paired with five echo locations (shown by open triangles). Solid circles represent localizations on trials in which the subject reported a single image; open circles, localizations of multiple images. Clearly, localizations of fused images were moved toward the azimuths of the echoes. What is more, the longer delays produced by echoic surfaces at azimuths or distances quite different from those of the source increased the reports of multiple images.

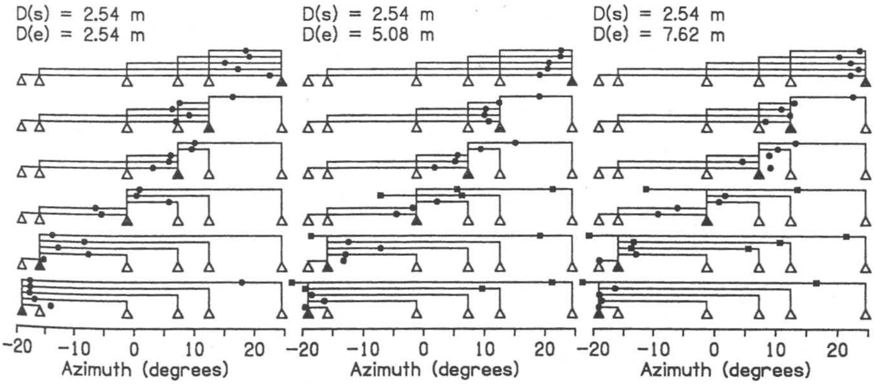

Fig. 4. Localizations of signals plus echoes from subject EJ. Triangles show the apparent azimuths of sources (solid) and echoes (open) *alone*, both corrected in accordance with biases shown in Fig. 3. Apparent locations *in the presence of echoes* are depicted by circles (single images) and squares (multiple images).

DISCUSSION

In discussing these data, we should first reiterate that problems can arise when there is no low-frequency energy, even with reasonably wide band stimuli. The biases shown in Fig. 3 agree with the Wightman *et al.* (1988) contention that localization of wide band stimuli is heavily dependent on IDTs in the low-frequency portions of a signal.

The main point, however, is that precedence requires a more complicated story than just echo suppression *or* disregard. It is clear that echoes affected both the discrimination of azimuth and the localization of single transients, but it is equally clear that localization was most strongly influenced by the azimuth of the source. It would seem that a softer definition of precedence that says that perceived direction is *near* the source is not far wrong.

Where to go next: It has taken us a long time to develop a simulated multi-source environment with good control over stimuli and responses, but the possibilities opened by this computer-controlled environment are great. At the moment, our sights remain on settling the primary issues raised at the beginning of this paper, those being an understanding of the nature of precedence, its level of processing and the tradeoffs in spatial hearing between a hardwired binaural system and more flexible cognitive processes. The data obtained so far have not fully answered those questions, and the work currently underway enlarges our focus by examining cases in which signals from the source and echo overlap in time, especially in situations with other, unrelated acoustic sources in the room. Similarly, we are studying the effects of modifying the echo spectrum by changing reflective properties of the simulated echoic surfaces. This is an important factor for understanding how echoes are distinguished from other extraneous sources, and a long-range goal is to examine localization in acoustically busy situations, where multiple independent signals are separated by acoustic *and* visual cues, i.e., "cocktail party" environments. In that regard, our technology continues to improve and will soon be enhanced by the addition of a large-scale visual display projected on the curtain that hides the speakers.

REFERENCES

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

Butler, R. A., Humanski, R. A. and Musicant, A. D. (1990). Binaural and monaural localization of sound in two-dimensional space. *Perception, 19*, 241-256.

Clifton, R.K. (1987). Breakdown of echo suppression in the precedence effect. *J. Acoust. Soc. Am., 82*, 1834-1835(L).

Gaskell, H. (1983). The precedence effect. *Hearing Res., 11*, 277-303.

Green, D. M. (1988). *Profile analysis: Auditory intensity discrimination*, Oxford University Press, New York.

Hafter, E. R. and Briolle, F. (1989). Localization of Complex Sounds in the Presence of Echoes. In: *Proc. 13th Int. Cong. on Acoust.*, pp. 495-498.

Hafter, E. R, and Buell, T. N. (1990). Restarting the Adapted Binaural System. *J. Acoust. Soc. Am., 88*, 806-812.

Hafter, E. R., Buell, T. N., Basiji, D. A. and Shriberg, E. E. (1988). Discrimination of direction for complex sounds presented in the free-field. In: *Basic Issues in Hearing* (H. Duifhuis, J. W. Horst and H. P. Wit, Eds.), Academic Press: London.

Hafter, E. R., Buell, T. N. and Richards, V. M. (1988). Onset-coding in lateralization: Its form, site and function. In: *Auditory Function: Neurobiological Bases of Hearing* (G. M. Edelman, W. E. Gall and W. M. Cowan, Eds.), Wiley: New York, N.Y.

Hartmann, W. M. and Rakerd, B. (1989). Localization of sound in rooms IV: The Franssen effect. *J. Acoust. Soc. Am., 86*, 1366-1373.

Middlebrooks, J. C., Makous, J. C. and Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Amer., 86*, 89-108.

Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. Am., 30*, 237-246.

Perrott, D. R., Marlborough, K., Merrill, P. and Strybel, T. Z. Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate. *J. Acoust. Soc. Amer., 85*, 282-288.

Poesselt, C., Schroeter, J., Optiz, M., Divenyi, P. and Blauert, J. (1986). Generation of binaural signals for research and home entertainment. In: *Proc. 12th Int. Cong. Acoust.*, B1-6.

Rakard, B. and Hartmann, W. M. (1985). Localization of sounds in rooms, II: The effects of a single echoic surface. *J. Acoust. Soc. Am.*, *78*, 524-533.

Saberi, K. and Perrott, D. R. (1990). Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate. *J. Acoust. Soc. Am.*, *87*, 1732-1737.

Wallach, H., Newman, E. B. and Rosenzweig, M. R. (1949). The precedence effect in sound localization. *Am. J. Psychol.*, *57*, 315-336.

Wenzel, E. M. (1991). Three-dimensional virtual acoustic displays. *NASA TM 103835*, to appear in *Multimedia and Multimodal Interface Design* (M. Blattner and R. Dannenberg, Eds.), ACM Press, Addison-Wesley: Reading, MA.

Wightman, F. L. (1991). Personal Communication.

Wightman, F. L. and Kistler, D. J. (1989). Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoust. Soc. Am.*, *85*, 858-867.

Zurek, P. M. (1987). The precedence effect. In: *Directional Hearing* (W. A. Yost, and G. Gourevitch, Eds.), Springer-Verlag: New York.

Commentaries

SCHARF

When clicks appeared to recede into the distance, was there any evidence of loudness constancy ?

HAFTER

You are referring to a statement made during the talk in which I refer to some of our preliminary observations of the effects of echoes on the sense of auditory distance. Thus far, we have simulated only the situation in which a single click moves towards or away from the listener along the median plane. When there is no echo, these stimuli generally sound like a click that waxes and wanes in level. However, when we use multiple speakers to simulate a wall running parallel to the median plane, the signal-plus-echoes appear to move in space, looming or receding along the median azimuth. Now, in response to your question, our informal observations are that in these cases, loudness constancy does seem to hold.