Echo suppression in the horizontal and median sagittal planes

Brad Rakerd
Department of Audiology and Speech Sciences, Michigan State University, East Lansing, Michigan 48824

William M. Hartmann
Department of Physics, Michigan State University, East Lansing, Michigan 48824

Joy Hsu[a]
Polytechnic School, Pasadena, California 91106

(Received 25 January 1999; revised 13 September 1999; accepted 30 October 1999)

Experiments were performed to measure two kinds of suppression threshold for running speech: echo threshold, defined here as the minimum level at which it was possible to detect that an echo was present, and masked threshold, defined as the minimum level at which it was possible to detect that a lagging sound was present at all. Both thresholds were measured using a geometry in which sound sources and reflections were distributed over the horizontal plane (left, front, and right locations) and a geometry in which they were distributed over the median sagittal plane (front, overhead, and rear locations). The predominant sound localization cues are different for these two geometries, and the experiments measured the consequences of this difference for suppression. Echo thresholds were found to have a comparable dependence on the delay of the lagging sound for the horizontal and median sagittal planes. Masked thresholds, which were systematically 8–15 dB lower than echo thresholds, also showed a comparable dependence on delay for the two planes. Overall, these results support the idea that echo suppression is functionally similar whether locations are cued by interaural differences in time and intensity, or by spectral features introduced by the head-related transfer function. [S0001-4966(00)02202-5]

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

INTRODUCTION

In a room, a sound comes to a listener many times over. It arrives first direct from its source, and thereafter from all around as the sound reflects and re-reflects from room surfaces. The reflected copies are a potential distraction for the listener, but they rarely distract to any great extent because a remarkable neural process suppresses their perception as echoes (Blauert, 1983). The present paper is motivated by a recent finding regarding the neurophysiology of echo suppression (Litovsky and Yin, 1994; Yin and Litovsky, 1994; Litovsky *et al.*, 1997). Recordings made in the inferior colliculus of cat point to an equivalence of suppression for sound sources and reflections spatially separated from one another in the horizontal plane (HP) and sources and reflections separated in the median sagittal plane (MSP). This comparison is interesting because the predominant directional cues are different for the two planes. In the HP, interaural difference cues are paramount (Mills, 1958; Durlach and Colburn, 1978). In the MSP, the most important cues are spectral shape cues introduced by the ears, head, and upper torso, which act as acoustical filters with different frequency responses for different angles of sound incidence (Shaw and Teranishi, 1968; Hebrank and Wright, 1974; Searle *et al.*, 1976; Middlebrooks and Green, 1991). Evidence of equivalence for the HP and the MSP argues for a suppression mechanism that is realized in the nervous system at a level where the locations of sources and reflections are represented abstractly, indifferent to the particulars of their spatial cuing.

The present study made a psychophysical comparison of echo suppression in the HP and the MSP, in complement to the neurophysiological comparison. It focused on human listeners' echo suppression for speech, which lasts for tens of milliseconds (Haas, 1951; Lochner and Burger, 1958), a time constant of the same order as that of the suppression that has been measured neurophysiologically. There were two experiments.

I. EXPERIMENT 1—ECHO THRESHOLD

Experiment 1 measured listeners' *echo threshold* for connected speech, defined here as the level at which a delayed copy of the speech was just barely audible as an image distinct from the direct sound. Measurements were made in the HP and the MSP, at delay times of 20, 40, 60, and 80 ms. The speech samples were newspaper articles, read aloud in an anechoic room by a female talker and by a male talker. Forty-five minutes of their speech were recorded onto digital audio tape and played back as needed for the experiment.

Five subjects participated in the study, S1, S2, and S3 were young adult listeners (two females, one male; ages 17 to 20 years) with normal hearing thresholds and no prior experience in psychophysical listening studies. S4 and S5 were older listeners (both males, ages 45 and 59), with some high-frequency hearing loss. S4 and S5 were experienced listeners. Subjects S1, S4, and S5 were authors.

The subjects were tested individually, in a 3.0 m (wide) ×4.3 m (long)×2.4 m (high) anechoic room (IAC #107840). During testing, a subject sat still, facing straight ahead. A metal guidebar rested atop the subject's head and helped the subject maintain a fixed head position. For the HP condition, three loudspeakers were placed at the height of the subject's

[a]Present address: Stanford University, Stanford, CA 94305.

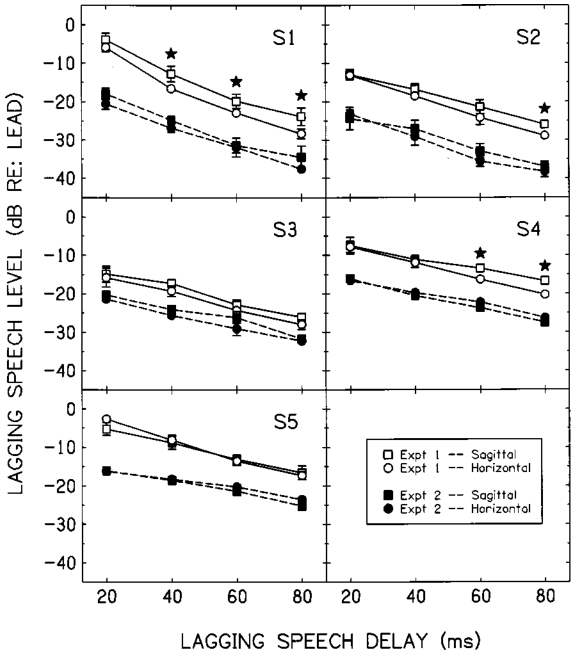

FIG. 1. Open symbols connected by solid lines show echo thresholds for running speech measured in experiment 1. Filled symbols connected by dashed lines show masked thresholds measured in experiment 2. Results are given separately for each subject (different figure panels). Thresholds (in dB *re*: the level of the lead) are plotted as a function of the delay time between leading and lagging copies of the speech. Plots with circles are for sources in the horizontal plane; plots with squares are for sources in the median sagittal plane. Error bars show ±1 standard deviation over test runs. Stars highlight instances in which error bars for the two planes do not overlap.

ears, and 120 cm away. One speaker was directly in front of the subject; the others were 90 deg off to the left and right. For the MSP condition, loudspeakers were placed directly in front, directly above, and directly behind the subject's head.

Test runs were done separately for the HP and the MSP. On each test trial, we directed a leading copy of the speech to one loudspeaker in the plane under test, and a lagging copy to the same speaker or to a different speaker in that plane. The subject was given control over the level of the lagging copy. At the highest level, which was 10 dB above that of the lead, the lagging copy was audible as an echo, distinct from the direct sound. The subject was instructed to reduce the level down to the point where the echo was as faint as possible while still remaining audible. Across trials, the leading and lagging loudspeakers were randomly varied, as was the delay time of the lag. One complete test run included all possible combinations of these variables (3 lead locations × 3 lag locations × 4 delay times = 36 trials). Subjects typically took about 20 min to complete a run. They did a total of four runs in the HP condition and four in the MSP condition. The order of these runs was random and different for each subject.

The results of experiment 1 are given by the open-symbol plots in Fig. 1. Echo thresholds (measured in dB *re*: the level of the leading sound) are plotted separately for each subject (different figure panels) and for the HP (open circles) and the MSP (open squares). Each plot gives a subject's mean threshold as a function of the lagging speech delay time. Error bars (in some cases smaller than the plotting symbols) indicate ± one standard deviation over the four runs that a subject did in each plane. All five subjects performed similarly in the HP and the MSP. For both planes, there was substantial suppression at the 20-ms delay time and progressively decreasing suppression with increasing delay times. There was a marked linear trend to the results. Best-fit lines (which accounted for over 90% of the variance in both planes) had nearly equal mean slopes for the HP (−0.25 dB/ms) and the MSP (−0.23 dB/ms).

For two subjects (S3 and S5), echo threshold levels for the HP and the MSP were close together at every value of the delay, never differing by more than the standard deviation over runs shown by the error bars. For the other three subjects (S1, S2, and S4), HP and MSP thresholds did diverge by amounts greater than the standard deviation at one or more values of the delay. Stars in Fig. 1 highlight the instances where this was the case (six in all). In every starred case, the HP echo threshold was lower than the corresponding MSP threshold. This effect was even present in the results for the one location that was common to both planes (center in the HP, front in the MSP).[1] Given that fact, and the fact that test runs were done separately for the HP and the MSP, we conclude that the plane difference reflects a context-sensitive shift in the subjects' criteria for decision making about echo thresholds. Just why this shift consistently favored slightly lower thresholds for the HP is unclear.

II. EXPERIMENT 2—MASKED THRESHOLD

Even after the level of a lagging sound is reduced to the point where an echo is no longer perceptible, audible loudness and coloration effects remain. Experiment 2 determined how much additional attenuation would be required to eliminate all audible effects of the lagging sound. The design of this *masked threshold* experiment was the same as for experiment 1, but the instructions to the subjects were different. Subjects were instructed to reduce the level of the lagging speech until they could barely detect that it was present at all. To aid in finding this point, subjects were given a pushbutton to press to remove the lagging sound altogether. On each test trial, they switched the lagging sound in and out and adjusted its level for as long as needed to find the masked threshold point. The subjects of this experiment were the same as for experiment 1. They completed all testing for that experiment before undertaking this one.

The results of experiment 2 are given by the filled-symbol plots in Fig. 1. Masked thresholds for all five subjects were 8 to 15 dB below the corresponding echo thresholds, with similar functions for the HP and the MSP. Best fit lines for the masked threshold data had identical mean slopes of −0.20 dB/ms for the HP and the MSP. Masked thresholds for the HP and the MSP were never further apart than the error bars for any subject at any delay time.

III. DATA ANALYSIS

A. Correlations

Figure 1 provides visible evidence that the subjects performed similarly in both experiments and in both planes. A correlation analysis was done to get a measure of the

TABLE I. Correlations among echo thresholds measured in experiment 1 and masked thresholds measured in experiment 2 for the horizontal plane (HP) and the median sagittal plane (MSP).

	Echo threshold		Masked threshold	
	HP	MSP	HP	MSP
Echo Threshold				
HP	...			
MSP	0.98	...		
Masked Threshold				
HP	0.93	0.89	...	
MSP	0.95	0.93	0.98	...

strengths of those relationships. Correlations were computed over the 20 values for each experiment/plane representing thresholds for each of the five subjects measured at each of the four delay times. Table I shows the correlation matrix. All of the correlations were high (0.89 or above), indicating a parallel dependence on lagging speech delay time throughout. Two correlations speak directly to the HP–MSP comparison that is at the heart of the present study. The correlation between HP and MSP echo thresholds measured in experiment 1 was 0.98. The correlation between HP and MSP masked thresholds measured in experiment 2 was 0.98 as well.

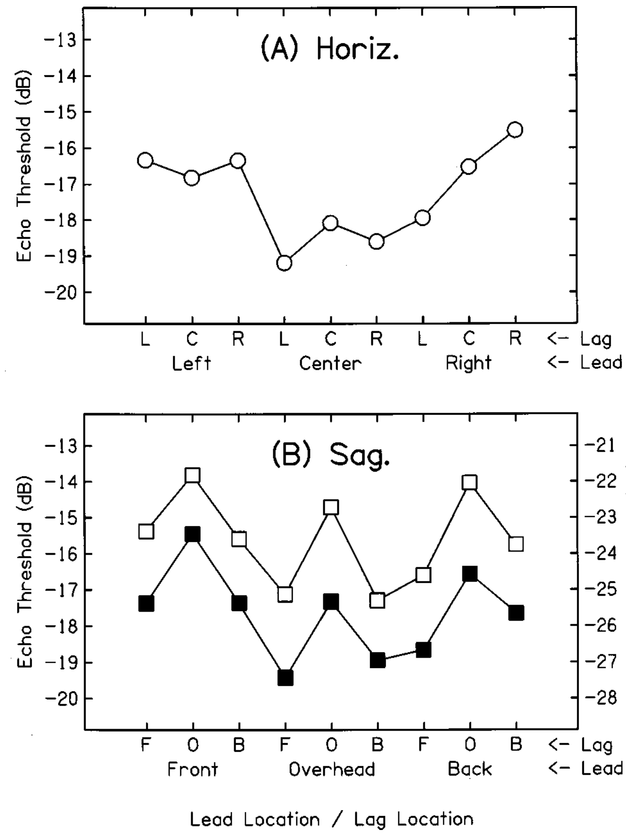

FIG. 2. Threshold results (in dB *re*: the level of the lead) averaged over all five subjects and four lagging delay times, and plotted a function of the different loudspeaker combinations used to present leading and lagging copies of the speech. (A): Echo thresholds (open circles) for loudspeakers arrayed in the horizontal plane (left, center, right locations). (Masked threshold, not plotted, showed no significant differences in this plane; see text.) (B): Echo (open squares) and masked (filled squares) thresholds for loudspeakers in the median sagittal plane (front, overhead, back locations).

B. Individual source locations

Over the course of an experimental run, stimuli were presented from nine different combinations of leading and lagging loudspeakers. Statistical comparisons showed that there were significant differences among the nine for echo threshold results in the HP [$F(8,32)=2.27$; $p<0.05$] and for both echo and masked threshold results in the MSP [echo: $F(8,32)=7.30$; $p<0.001$; masked: $F(8,32)=8.59$; $p<0.001$]. There was no significant difference among the loudspeaker combinations for the masked threshold results in the HP ($p>0.05$). All of the loudspeaker-combination effects were statistically independent of the delay-time factor that was of chief interest in this study (no significant interaction with delay in any analysis; $p>0.05$).

Figure 2(A) plots the echo threshold results for the HP as a function of the different lead-lag loudspeaker combinations. The plot shows that thresholds for loudspeaker combinations that were in the same location (e.g., lead left, lag left) were comparable to thresholds for combinations that were in different locations (e.g., lead left, lag right). We found this result throughout. There was no statistical difference between same-location and different-location combinations in any analysis ($p>0.05$). Yang and Grantham (1997) also found no consistent effect of loudspeaker separation (including no separation) on echo thresholds measured in the HP with click stimuli.

Figure 2(A) shows that echo thresholds were uniformly lower when the lead speaker was at the center than when it was at the left or right. (Thresholds for the center-left, center-center, and center-right combinations were the three lowest in the set). Seraphim (1961) also noted a disparity in echo suppression for speech between center and side locations in the horizontal plane. In that instance, a lagging speech sound at the center was found to be suppressed more easily than a sound off to the side. Here the center location proved less effective than the side locations as an echo suppressor.

Figure 2(B) gives lead-lag plots of echo and masked thresholds for the MSP. Both plots show a generalized weakness for the overhead location, relative to front and back. When leading, overhead weakly suppressed reflections coming from the other two locations (over-front and over-back had the lowest thresholds in the set of nine). When lagging, overhead was itself readily suppressed by leads from front and back (front-over and back-over had the highest thresholds). A substantial majority of the variation in both the echo and masked thresholds could be attributed to this result, which amounted to an effective 2-dB reduction in the strength of the overhead source in the MSP. Results for the front location and the back location were comparable overall for both echo thresholds and masked thresholds.

IV. SUMMARY AND CONCLUSIONS

When listening to speech in a room, a listener is rarely aware of acoustical reflections or reverberation. This is because the listener's brain suppresses the perception of speech echoes for tens of milliseconds after the arrival of a direct sound (Haas, 1951; Lochner and Burger, 1958). In two experiments, we measured thresholds associated with this effect for sound sources and reflections distributed in the horizontal plane and in the median sagittal plane.

Experiment 1 measured a listener's echo threshold—the level at which an echo was barely audible—as a function of the delay time between leading and lagging copies of running speech. Results were overall similar for the horizontal and median sagittal planes, particularly with respect to their delay-time dependence. One difference between the planes was that measured echo thresholds were slightly lower for the horizontal plane than for the median sagittal plane in several instances. We tentatively attributed that result to response bias. Experiment 2 measured a listener's masked threshold—the level at which any effect of lagging speech was barely audible—again, as a function of delay time. The results were comparable for the HP and the MSP in every important respect. Finally, a correlation analysis showed a high level of similarity between the HP and MSP results for both the echo ($r=0.98$) and masked ($r=0.98$) threshold experiments.

The present psychophysical findings of equivalence in echo suppression for the HP and the MSP can be interpreted two ways. It is possible that the suppression occurs at a neural processing site that is indifferent to source location. Alternatively, the findings may be seen as consistent with recent neurophysiological evidence (Litovsky and Yin, 1994; Yin and Litovsky, 1994; Litovsky *et al.*, 1997) for an echo suppression mechanism mediated by higher auditory centers where binaural and spectral cues to location are combined.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIDCD) Grant No. DC00181. The National Science Foundation provided support for Joy Hsu, through the Michigan State University High School Honors Science Program (Grant No. ESI-94-52753). We thank David Perrott for many useful comments on a previous version of this article.

[1] One-ninth of all the trials for the two planes were identical, with leading and lagging speech sounds coming from directly ahead of the listener in both cases (center location in the HP, front location in the MSP). An analysis of these trials showed the same disparity between the two planes that was shown for the larger set of stimuli (consistently lower echo thresholds for the HP).

Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

Durlach, N. I., and Colburn, H.S. (**1978**). ''Binaural phenomena,'' in *Handbook of Perception*, Vol. 4, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 365–466.

Haas, H. (**1951**), ''Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache [On the influence of a single echo on the intelligibility of speech],'' Acustica **1**, 49–58.

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.

Litovsky, R. Y., and Yin, T. C. T. (**1994**). ''Physiological correlates of the precedence effect: free-field recordings in the inferior colliculus of the cat,'' Assoc. Res. Otolaryngol. **85**, 337 (Abstract).

Litovsky, R. Y., Rakerd, B., Yin, T. C. T., and Hartman, W. M. (**1997**). ''Psychophysical and physiological evidence for a precedence effect in the median sagittal plane,'' J. Neurophysiol. **77**, 2223–2226.

Lochner, J. P. A., and Burger, J. F. (**1958**). ''The subjective masking of short time delayed echoes, their primary sounds, and their contributions to the intelligibility of speech,'' Acustica **8**, 1–10.

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.

Mills, A. W. (**1958**). ''On the minimum audible angle,'' J. Acoust. Soc. Am. **30**, 237–346.

Searle, C. L., Braida, L. D., Cuddy, D. R., Davis, M. F., and Colburn, H. S. (**1976**). ''Model for auditory localization,'' J. Acoust. Soc. Am. **60**, 1164–1175.

Seraphim, H. P. (**1961**). ''Über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachshall [On the perception of multiple reflections of speech sounds],'' Acustica **11**, 80–91.

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. **78**, 524–533.

Yang, X., and Grantham, W. (**1977**). ''Echo suppression and discrimination aspects of the precedence effect,'' Percept. Psychophys. **59**, 1108–1117.

Yin, T. C. T., and Litovsky, R. (**1994**). ''Physiological studies of the precedence effect in the inferior colliculus of the cat,'' in *Advances in Hearing Research: Proceedings of the 10th International Symposium on Hearing,* edited by G. A. Manley, G. M. Klump, C. Koppl, and H. Oeckinghaus (World Scientific, Singapore).