Abstract
The Hearing-in-Noise Test (HINT) is able to measure the benefit to speech intelligibility in noise conferred when the noise masker is displaced 90 degrees in eccentricity from a speech source located at zero degrees azimuth. Both psychoacoustic and neurophysiological data suggest that the perceptual benefit of the 90-degree azimuth separation would be greatest if the speech and noise were presented in different acoustic hemifields, and would be smallest if the two sources were in the same acoustic hemifield. The present study tested this hypothesis directly in ten normal-hearing adult listeners. Using the HINT stimuli, we confirmed the hypothesis. Release from masking scores averaged 8.61 dB for “between-hemifield” conditions, 6.05 dB for HINT conditions, and 1.27 dB for “within-hemifield” conditions, even though all stimulus configurations retained a 90-degree angular separation of speech and noise. These data indicate that absolute separation of speech and noise alone is insufficient to guarantee a significant release from masking, and they suggest that what matters is the location of the stimulus elements relative to the left and right spatial perceptual channels.

Key Words: Acoustic hemifield, Hearing-in-Noise Test, noise, spatial masking, speech reception threshold

Abbreviations: ANOVA = analysis of variance; HINT = Hearing-in-Noise Test
One of the most troubling tasks for the hearing-impaired listener, and therefore one of the potentially valuable outcome measures for audiological or otological interventions, is the listener’s ability to comprehend speech in noisy environments. There is a long history of research that has attempted to identify spectral or temporal processing impairments in listeners with hearing loss, and the consequences of those impairments for the intelligibility of speech in the presence of noise backgrounds (e.g., Elliott, 1975; Arlinger and Dryselius, 1990; Dubno and Dirks, 1990; Festen and Plomp, 1990; Schorn and Zwicker, 1990; Phillips et al, 1994; Stuart and Phillips, 1996). Recently, there has been interest in the effects specifically of the spatial separation of speech and background noise on speech intelligibility (Yost et al, 1996; Noble et al, 1997; Peissig and Kollmeier, 1997; Kidd et al, 1998).

A “Hearing-in-Noise Test” has been developed (HINT: Nilsson et al, 1994; Soli and Nilsson, 1994). This test presents highly standardized sentences (Nilsson et al, 1994) from a source at zero eccentricity and measures speech reception thresholds (SRTs) in three conditions distinguished by the location of concurrent speech-spectrum masking noise (0 [“full mask”], -90, and +90 degrees azimuth: Soli and Nilsson, 1994). The measure of interest is the perceptual benefit (“release from masking”) obtained from having a 90-degree spatial separation of speech and masker, and this benefit is typically on the order of about 7 dB at speech reception threshold (SRT) in normal listeners (Soli and Nilsson, 1994).

Spatial separation of speech and noise in the sound field could facilitate recovery of the speech signal through at least two mechanisms. Spatial separation will provide a better signal-to-noise ratio at the ear closest to the speech, particularly for those high-frequency spectral components that are more susceptible to shadowing by the head and pinnae (after Hirsh [1950]; Saberi et al [1991]). Spatial separation of speech and noise also offers the opportunity for the two sounds to activate different populations of central binaural neurons, since central neurons are often differentially activated by interaural disparities in stimulus timing and level (after Phillips and Irvine, 1981; Phillips and Brugge, 1985; Zurek, 1993). In both cases, differential encoding of the speech signal would be optimized when speech and noise are located in opposite acoustic hemifields.

Without necessarily addressing the question of mechanism directly, at least four lines of evidence suggest that the central auditory system uses a broad, hemifield framework for the processing of spatial information. One is that the spatial release from masking of one broadband sound by another is greater when the two sources span the midline than when they do not (Saberi et al, 1991). Second, auditory temporal gap detection for markers spanning the midline is more impaired than is gap detection for markers with the same spatial separation but within a single hemifield (Boehnke and Phillips, 1999). Third, central neurons with spatially restricted receptive fields usually have hemifield-shaped receptive fields whose medial borders are within about 20–30 degrees of the midline (e.g., Middlebrooks and Pettigrew, 1981; Phillips and Brugge, 1985; Imig et al, 1990). Finally, following unilateral auditory cortical damage, sound localization behavior is impaired for the auditory hemifield contralateral to the damage (Sanchez-Longo and Forster, 1957; Jenkins and Masterton, 1982; Kavanagh and Kelly, 1987).

The hemifield architecture implied by the foregoing suggests that the greatest
perceptual benefit of a 90-degree separation of speech and noise should therefore occur when the two stimulus components fall into different acoustic hemifields, and the least perceptual benefit of a 90-degree separation should occur when the two stimuli fall within a single hemifield. The HINT in its current form (Soli and Nilsson, 1994) does not test either of these stimulus configurations; rather, it has one stimulus located deep within a hemifield and the other at the boundary between left and right hemifields. Release from masking values obtained from the HINT are therefore likely to fall between those that would be obtained using purely within- and between-hemifield stimulus configurations.

The purpose of the present study was to test this hypothesis directly in ten normal-hearing listeners, using the HINT test stimuli. The general strategy was to compare release from masking scores for 90-degree separation conditions that differed in their absolute eccentricities. We compared performance for conditions in which speech and noise fell in the same hemisphere (left or right, and with speech in front or behind), different hemispheres (45 degrees either side of the midline, with speech on the left or right), and in the HINT configuration (with speech at zero degrees azimuth, and noise on the left or right). The data obtained constitute a new expression of the hemifield architecture of human spatial processing mechanisms (after Boehnke and Phillips, 1999).

METHODS

Subjects were 10 young adults (6 males, 20–30 years of age, mean = 24 years), all of whom had audiometric thresholds better than 15 dB HL at octave frequencies from 0.25 to 8.0 kHz. None of the participants had prior experience in psychoacoustical studies. All procedures used in this study received ethical approval by the Dalhousie University Research Ethics Board.

The 220 standardized sentences and speech spectrum noise from the HINT compact disc (© House Ear Institute, 1994) were saved as individual wave (.wav) files and stored digitally on the hard drive of a Macintosh G3 computer. On each stimulus trial, a single sentence was randomly selected (without replacement) and presented with or without accompanying speech-spectrum noise. On trials in which both speech and noise were presented, the noise was a randomly selected piece of the HINT speech-spectrum noise. The noise was shaped with 20 msec rise and fall times, and it began (and ended) 500 msec before (and after) the HINT sentence. The digital signals in MATLAB were transformed to analog through an ECHO GINA-24 multichannel sound card. Activated channels were amplified independently and then transduced by one of eight matched Optimus PROx44 two-way speakers (see below).

The listener was seated, with his or her head resting on a chinrest (to align the head with zero degrees azimuth and to prevent leaning), in the center of an Eckel sound-attenuating room whose internal walls and ceiling had been lined with two-inch Auralex wedges, and whose floor was carpeted. The internal room dimensions were 95 inches deep x 83 inches x 80 inches high, and the chinrest height was 40 inches. The listener was fitted with a headset microphone that left the ears unobstructed, and his task on each trial was to repeat the spoken sentence into the microphone. Outside the sound booth, the experimenter scored the subject's response as correct or incorrect according to the HINT's instruction manual.

All subjects were tested individually. The subject was surrounded by eight matched, equally spaced, Optimus PROx44 two-way speakers located straight ahead (0 degrees azimuth) and at +/- 45 degrees, +/- 90 degrees, +/- 135 degrees and 180 degrees, where “+” and “-” denote Left (L) and Right (R) auditory hemifields respectively. The diameter of the speaker array was 72 inches. The sound pressure level of the noise stimulus was fixed at 59 dB (A weighted, Brül & Kjaer sound level meter model 2209; measured with a microphone located where the center of the listener’s head would otherwise have been).

There were ten stimulus conditions tested in each subject. The first four were the original HINT conditions: speech alone at zero degrees, speech and noise both at zero degrees (“full mask”), speech from zero degrees and noise from +/- 90 degrees. Two further conditions were the “between-hemifield” ones: speech from -45 degrees and noise from +45 degrees, and the spatial mirror image of that configuration. The remaining four conditions were “within-hemifield” ones: speech at -45 and noise at -135 degrees (and their spatial mirror image), and speech at +45 and noise at +135 degrees (and their spatial
mirror image). Thus, all conditions in which both speech and noise were presented preserved the 90-degree angular separation of speech and noise but varied in the absolute eccentricities spanned by that separation.

The general strategy was to obtain the SRT for every one of the ten conditions. To do this, we used ten interleaved staircases of speech stimulus level, one for each of the ten conditions, concurrently. The staircases were of the two-down/one-up kind (after Levitt, 1971), with the exception that at the beginning of each staircase, for every correct response, speech stimulus level was reduced by the adaptive step until the first error. The size of the adaptive step was 2 dB until the third reversal, after which the adaptive step was 1 dB in size. The speech level used at the beginning of each staircase was chosen by pilot testing to be about two to three trials above SRT. The interleaved staircases continued until the full complement of 220 HINT sentences had been exhausted. This design ensured that no listener in the study heard the same sentence twice. In practice, the design resulted in staircases containing at least four reversals in the sign of the adaptive step. SRT was defined as the mean of the speech levels at which the adaptive step changed sign.

Data collected were SRTs for each spatial configuration condition. These values were then subtracted from the SRT for the full-mask condition, to provide a “release from masking” score. Note that for this analysis to be meaningful, we had first to show that the full-mask SRT (with speech and noise at zero degrees azimuth) was equivalent to that for full-mask conditions at every azimuth tested. This was confirmed empirically in pilot testing.

**RESULTS**

Data collected were speech levels required for 70.7 percent correct sentence repetition for each speech/noise spatial condition. These values were expressed as release-from-masking scores by subtracting the threshold at each condition from the listener’s threshold at full-mask condition; that is, speech and noise presented from zero degrees azimuth. The release from full-masked threshold was determined for ten listeners for each of the eight test conditions. Figure 1 shows the mean release-from-masking scores for each of the eight spatial separation conditions; variance in the data is indicated by standard error bars. The diagrams at the top of Figure 1 are schematics depicting the speaker locations used to obtain the measurements depicted by the pairs of histograms below the schematics.

There was more release from full-masked threshold for the between-hemifield conditions than for the within-hemifield conditions. That is, the 90-degree separation of speech and noise provided greater release from masking when the speech signal and noise were dispersed across the midline (i.e., speech -45 degrees, noise 45 degrees) than when the speech and noise were located within the same acoustic hemifield (i.e., speech at 45 degrees, noise at 135 degrees). The original HINT conditions, in which the speech was presented at zero degrees and the noise was located 90 degrees to the left or right along the azimuth, usually resulted in intermediate release from masking. The grand mean release from masking scores (i.e., averaged across listeners and subconditions) were 1.27 dB for the within-hemifield conditions (white bars in Figure 1), 6.05 dB for the HINT conditions (hatched bars in Figure 1), and 8.60 dB for the between-hemifield conditions (black bars in Figure 1).

**Figure 1** Mean release from masking scores, averaged over the ten listeners, shown as a function of stimulus condition. Upper schematics show the spatial locations of the speech and noise stimuli for the stimulus conditions associated with the data histograms below them. Black bars represent mean scores for the between-hemifield conditions; hatched bars represent data from the HINT conditions; and open bars represent data from the within-hemifield conditions. L, R, F, and B refer to whether the speech was presented on the Left, Right, Front, or Back position relative to the noise. Bars indicate standard errors.
The data in Figure 1 are awkward to analyze statistically because the overall experimental design is not ideal for an omnibus analysis of variance (ANOVA). However, statistically appropriate contrasts can be made between the data for certain subsets of conditions, and these contrasts directly answer the questions posed by our study. One ANOVA was used to analyze the difference between data from the HINT and the between-hemifield conditions. A within-subjects, 2 (Condition) by 2 (Signal Location: speech to the left or right of the noise) design was used. There was a main effect of Condition (F [1,9] = 18.2, p < 0.05), indicating that release from masking was greater when the speech and noise were spatially separated across the midline as opposed to being in the original HINT configurations. There was no main effect of Signal Location (F [1,9] = 4.35, p > 0.05), suggesting that no significant left (cerebral) hemisphere advantage for verbal material was evident. There was no significant interaction between the two variables (F [1,9] = 0.075, p > 0.05).

A second ANOVA was used to compare data from the two within-hemifield categories. This ANOVA used a within-subjects, 2 (speech front/back) by 2 (hemifield) design. There was no significant main effect of front/back condition (F [1,9] = 3.51, p > 0.05), or hemifield (F [1,9] = 0.19, p > 0.05), indicating that any left hemisphere advantage or head shadow effect was not significant. There was no interaction between these variables (F [1,9] = 4.51, p > 0.05).

The third comparison is of the data from the HINT and the within-hemifield conditions. This contrast is not amenable to an ANOVA, because the stimulus dimensions along which the two sets of conditions differ are not matched. Nevertheless, as might be expected from the data in Figure 1, for nine of the ten listeners, release from masking scores for the within-hemifield conditions were nonoverlapping with those for the HINT (and between-hemifield) conditions. Note that a direct consequence of this is that HINT conditions typically evoke a release from masking whose magnitude falls between those seen in the within-hemifield and between-hemifield conditions.

**DISCUSSION**

The purpose of this study was to provide evidence on whether the release from masking afforded by a 90-degree separation in azimuth of speech and noise varies with the disposition of the two sounds in relation to the left and right acoustic hemifields. This question was prompted by previous studies using nonspeech signals, which demonstrated that the interaction of two signals as a function of their spatial separation was dependent less on their absolute separation than on their azimuthal positions relative to the midline (Saberi et al, 1991; Boehnke and Phillips, 1999). Specifically, spatial release from simultaneous masking and temporal gap detection between two markers is more affected by a given spatial separation across the midline than by the same separation within an auditory hemifield. The current form of the HINT usually employs a zero-degree azimuthal location for the speech signal, and +/- 90-degree eccentricities for the noise masker (Soli and Nilsson, 1994). If release from masking of speech follows the same pattern as release from masking of nonspeech signals (Saberi et al, 1991), then the current HINT should provide a release from masking intermediate in magnitude between that obtained with the same 90-degree separation of speech and masker, but for sources in the same or the opposite acoustic hemifields.

This expectation was confirmed (Figure 1). The release from masking we observed using the HINT spatial configurations averaged 6.05 dB. This is close to the 7.0 dB advantage reported by Soli and Nilsson (1994). The ~1 dB difference between the two reports likely has its origins in methodological factors. The present study employed completely randomized sentence presentation, and a formal adaptive tracking strategy for measurement of speech reception threshold. Using the present methodology but retaining the 90-degree separation of speech and masker, but for sources in the same or the opposite acoustic hemifields.

There are two related implications of these data. One concerns the phenomenology of spatial hearing. The dynamic range (i.e., the effect size) of the HINT could be enhanced by 40 percent from around 6 dB to about 8.6 dB, simply by using speaker locations in opposite acoustic hemifields. This is advantageous if the HINT were to be used specifically to probe the integrity of single listener’s ability to use spatial separation to
extract speech from a noise background: for a given measurement variability, the greater the independent variable’s effect size on the measurement, the more finely can the measurements be differentiated. The further implications of the present findings for management strategies in persons with hearing impairment are perhaps less clear. This is because while the present study has provided a means further improving the intelligibility of speech against a noise background, it does so by misaligning the head with the speech source. This places the listener in the predicament of having to retain visual contact with the speech source by using eye position rather than head orientation.

The second implication concerns how the present data inform us about the architecture of auditory spatial processing. Four previous lines of evidence point to a hemifield architecture of auditory spatial processing mechanisms. Two lines of evidence come from animals. Of central neurons that have spatially restricted receptive fields, the most common disposition of the receptive field is across the contralateral acoustic hemifield, with medial receptive field edges within about 20–30 degrees of the midline (Middlebrooks and Pettigrew, 1981; Moore et al, 1984; Phillips and Brugge, 1985; Imig et al, 1990). This means that central neurons having those receptive fields form neural “channels” for representation of the contralateral auditory hemifield. Second, unilateral brain lesions result in sound localization deficits for sources only in the contralateral acoustic hemifield (Jenkins and Masterton, 1982; Kavanagh and Kelly, 1987).

A third thread of research is on human spatial release from masking (see Saberi et al, 1991). For a nonspeech signal located in one hemifield, the effect of a concurrent wideband masker of a given eccentricity difference is much greater if the two stimuli are in the same hemifield than if they are disposed across the midline. This is true even under monaural listening conditions, and so likely reflects the signal to noise ratio benefit conferred at the ear in the same hemifield as the signal (see also Hirsh, 1950). Additional data attesting to superior speech intelligibility for target sources when simultaneous sources are disposed on opposite side of the midline can be recovered from other experimental paradigms (Cherry, 1953; Yost et al, 1996; Kidd et al, 1998). Finally, temporal auditory gap detection thresholds for wideband markers are low when the sources are within one acoustic hemifield but higher when the markers are in opposite hemifields (Boehnke and Phillips, 1999). Whether this reflects poor acuity of between-channel timing operations in the spatial domain (Boehnke and Phillips, 1999) or some form of nonsimultaneous masking effects at either ear alone (after Oxenham, 2000) is unclear, but the hemifield “tuning” of the effect is what matters for the current argument. The present study has provided a new line of evidence that is compatible with all of the above: that the release from masking afforded to a speech signal by the signal’s spatial separation from a masker is greatest when the speech and noise are in opposite hemifields, and smallest when the speech and masker are in the same hemifield. The present data do not address the question of mechanism, but they do provide a new line of support for the notion that auditory spatial processing in man has a hemifield architecture.

Acknowledgment. This work was supported by grants from NSERC of Canada to D. P. Phillips, and from CLLRNet to D. P. Phillips and R. M. Klein. Special thanks are due to Dr. Manohar Bance and to Michel Comeau for helpful discussions that preceded this research, and to Kerry Walker, Gail Eskes, and three anonymous reviewers for comments on previous presentations of this work.

REFERENCES


