Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise

Trevor R. Agus[a)]
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow G31 2ER, United Kingdom; Department of Psychology, University of Strathclyde, 40 George Street, Glasgow G1 1QE, United Kingdom; and CNRS, Universite Paris Descartes, and Ecole Normale Superieure, 29 rue d'Ulm, 75005 Paris, France

Michael A. Akeroyd and Stuart Gatehouse[b)]
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow, G31 2ER, United Kingdom

David Warden
Department of Psychology, University of Strathclyde, 40 George Street, Glasgow G1 1QE, United Kingdom

(Received 8 October 2008; revised 24 July 2009; accepted 25 July 2009)

Three experiments measured the effects of age on informational masking of speech by competing speech. The experiments were designed to minimize the energetic contributions of the competing speech so that informational masking could be measured with no large corrections for energetic masking. Experiment 1 used a "speech-in-speech-in-noise" design, in which the competing speech was presented in noise at a signal-to-noise ratio (SNR) of −4 dB. This ensured that the noise primarily contributed the energetic masking but the competing speech contributed the informational masking. Equal amounts of informational masking (3 dB) were observed for young and elderly listeners, although less was found for hearing-impaired listeners. Experiment 2 tested a range of SNRs in this design and showed that informational masking increased with SNR up to about an SNR of −4 dB, but decreased thereafter. Experiment 3 further reduced the energetic contribution of the competing speech by filtering it into different frequency bands from the target speech. The elderly listeners again showed approximately the same amount of informational masking (4–5 dB), although some elderly listeners had particular difficulty understanding these stimuli in any condition. On the whole, these results suggest that young and elderly listeners were equally susceptible to informational masking.
 [DOI: 10.1121/1.3205403]

PACS number(s): 43.66.Dc, 43.66.Sr, 43.71.Lz, 43.71.Ky [RLF] Pages: 1926–1940

I. INTRODUCTION

Elderly listeners generally have more listening difficulties than young listeners, particularly when listening to one of several simultaneous talkers (CHABA, 1988). This undoubtedly stems in part from the reduced functioning of their ears, but there could also be some "higher-level" hearing difficulties, for example, in confusing the simultaneous talkers. Such difficulties have often been demonstrated experimentally in normal-hearing listeners, as there is a greater amount of masking when competing speech is present compared to some acoustically equivalent condition without speech (e.g., Brungart, 2001; Freyman *et al.*, 2001; Arbogast *et al.*, 2002). This effect is termed "informational masking." Previously, we have demonstrated that informational masking may be factor in people's self-report of their listening ability, in that an analysis of responses to the "Speech, Spatial, and Qualities of Hearing" questionnaire (Gatehouse and Noble, 2004) showed lower self-reported performance in situations likely to involve informational masking than in control situations (Agus *et al.*, 2009). Here, we report direct measurements of the effect of age and hearing loss on the informational masking of speech by speech.

How best to define informational masking remains an open question (e.g., Durlach *et al.*, 2003; Watson, 2005; Durlach, 2006), but in order to obtain an experimentally tractable definition it is convenient to use Lutfi's (1990) equation

$$\text{total masking} = \text{informational masking} + \text{energetic masking}, \qquad (1)$$

or, equivalently,

$$\text{informational masking} = \text{total masking} - \text{energetic masking}. \qquad (2)$$

Lutfi (1990; Oh and Lutfi, 1998) took (A) the masked threshold for a "stimulus known statistically" (i.e., a greater-uncertainty condition) as the measure of "total masking" in Eq. (2) and (B) the masked threshold observed for a "stimulus known exactly" (i.e., a minimal-uncertainty condition) as the measure of "energetic masking." The amount of uncertainty in a signal can be quantified for experiments with tonal signals, but it is not clear how it can be measured for speech signals. As the focus in this paper is on the informational

[a)]Author to whom correspondence should be addressed. Electronic mail: trevor.agus@ens.fr
[b)]Stuart Gatehouse was closely involved in the design and analysis of this experiment before his death in February 2007.

masking of speech by speech, we used corresponding definitions instead: (A) the measure of total masking was taken as the masked threshold in a condition with competing speech as part of the masker, and (B) the measure of energetic masking was taken as the masked threshold in an acoustically equivalent control condition *without* competing speech. That is,

$$\text{informational masking} = \text{SRT}_{\text{speech}} - \text{SRT}_{\text{nonspeech}}, \quad (3)$$

where $\text{SRT}_{\text{speech}}$ and $\text{SRT}_{\text{nonspeech}}$ are the speech-reception thresholds in the presence or absence of competing speech. This equation underlies much of the research on informational masking in speech, as it makes no claims about the source of the informational masking—it could be due to similarity or uncertainty (Watson, 2005), modulation masking (Kwon and Turner, 2001), "semantic interference" (Carhart *et al.*, 1969a), or otherwise unknown facets of speech-on-speech masking—although it does require that the speech and nonspeech maskers generate the *same* amount of energetic masking.

A modulated, speech-shaped noise is often used as the nonspeech masker (e.g., Festen and Plomp, 1990; Brungart, 2001), as it can represent both the overall temporal fluctuations of speech and its overall long-term spectrum. However, it still has *some* acoustical differences to speech—as indeed would any noise masker—and if these differences affect the energetic masking compared to actual speech, then the amount of informational masking derived using Eq. (3) could be misleading.

An alternative to finding a perfect acoustically equivalent masker is instead to use a masking noise in *both* conditions. That is, the masker in the total-masking condition is a combination of competing speech plus a noise, and the masker in the energetic-masking condition is just the noise. The essence of this design is that the noise should contribute almost all the energetic masking, while the competing speech contributes the informational masking but only a minimal amount of additional energetic masking: thus energetic and informational masking are conceptually separated. This goal can be achieved by taking advantage of the well-known result that the speech-reception threshold for a sentence in noise is generally negative—when expressed as a speech-to-noise ratio in decibels—meaning that the sentence has less power than the noise when it is at threshold (e.g., −8 dB; Hawkins and Stevens, 1950). For example, if a sentence at a long-term average speech-to-noise ratio of −8 dB is added to a noise at 60-dB SPL, then the sentence can (just) be identified, and it might be expected to create a small amount of informational masking. But because the level of the sentence (52-dB SPL) is considerably below the level of the noise (60-dB SPL), the total long-term average power of the combination is almost entirely due to the power of the noise and is indeed is less than a decibel greater (60.6-dB SPL).[1] The crucial experimental comparison is thus between the SRTs in the combined masker and in the noise-alone masker: if the difference is also 0.6 dB in this example, then the entire effect can be attributed solely to the average increase in acoustical power. If, however, the difference is larger than 0.6 dB, then there must have been some additional masking that cannot be attributed purely to the acoustical power of the maskers. Any such additional masking should mostly be informational masking, and so a simple application of Eq. (2) allows it to be calculated.

Carhart *et al.* (1968) pioneered this design. They used a speech-to-noise ratio of 0 dB and so assumed that the increase in masker power due to the combination was 3 dB; their noise masker was a modulated noise, not a static noise. As part of a larger experiment, they measured SRTs of spondees in either a combination of competing speech plus modulated noise or in modulated noise alone. They observed 7.8 dB more masking in the combined maskers: of this, 3 dB was attributed to the increase in total power of the maskers and so the remaining 4.8 dB was attributed to "semantic interference," which would now be termed informational masking. However, since modulated noise was used, some of the additional masking could be attributed to the speech "filling in" the dips in the noise, which would lead to an overall reduction in the modulations; this effect was estimated by Carhart *et al.* (1969a) to be approximately 3 dB. Carhart *et al.* (1969a) reported a similar experiment, differing in that it included unmodulated noise and a speech-to-noise ratio of approximately −3 dB; they observed 6.6 dB of additional masking beyond that expected due to differences in power.

The method of Carhart *et al.* (1968) was also used in one of two experiments that have studied the effect of age on informational masking. Tillman *et al.* (1973) found 3 dB of additional masking for younger listeners and 4 dB for the elderly listeners. When a second competing talker was added to the masker, elderly listeners showed 2 dB more informational masking than the young listeners. Taken together, these results suggest that elderly listeners may be more susceptible to the informational masking of speech, although only by about 1–2 dB. Another experiment that compared amounts of informational masking for young and elderly listeners is that of Li *et al.* (2004), who measured the *release* from informational masking, using a "precedence-effect" method similar to Freyman *et al.* (1999). They showed that elderly listeners had the same release from informational masking as younger listeners, and thus they inferred that the total amount of informational masking, before any was released, was the same for elderly and younger listeners.

Many studies have measured the effect of competing speech for hearing-impaired listeners (e.g., Tun and Wingfield, 1999; Tun *et al.*, 2002; Hornsby *et al.*, 2006; Humes *et al.*, 2006). These experiments were mostly not designed to allow a clear distinction between energetic masking and informational masking, however, but one important study by Arbogast *et al.* (2005) did explicitly measure informational masking. They used pure-tone vocoded sentences (Arbogast *et al.*, 2002), with the target speech and competing speech formed from different sets of modulated pure tones so that their mutual energetic masking was minimized. Arbogast *et al.* (2002) found that hearing-impaired listeners had much higher SRTs for the target speech in noise (which had the same long-term average spectrum as the pure-tone vocoded competing speech), and therefore less of their speech-on-speech masking was attributable to informational masking.

FIG. 1. A schematic illustration of the three stimulus conditions of experiment 1. The open black box represents the steady noise present in each condition. The text *Christmas is coming soon* represents the target speech, which starts 750 ms after the steady noise in each condition. In the SN condition (middle panel) the gray text represents the competing speech ("The dirty boy was washing the puppy licked his master"). In the MN condition (bottom panel) the gray waveform represents the modulated noise.

That is, their hearing-impaired listeners showed less informational masking than their normal-hearing listeners.

Taken together, these results are contrasting: those of Li *et al.* (2004) and Arbogast *et al.* (2005) imply that informational masking is no larger in elderly or hearing-impaired listeners than in younger normal-hearing listeners—and may even be less—but the results from Tillman *et al.* (1973) would suggest the opposite. The aim of the present experiments was therefore to compare the amount of informational masking observed in these groups and to further develop methodologies for comparing young and elderly listeners' susceptibility to informational masking of speech by speech. Experiments 1 and 2 used a "speech-in-speech-in-noise" design related to Carhart *et al.* (1968, 1969a), whereas experiment 3 used a design inspired by Arbogast *et al.* (2002, 2005). We attempted to increase the amount of informational masking obtained by having target- and competing-speech sentences that were similar to each other: they both used recordings from the same talker and were presented diotically. Thus the experiments omit some cues for distinguishing simultaneous sentences that would often be available in everyday conversation, such as talker differences and localization cues (Freyman *et al.*, 1999; Brungart, 2001). Unlike many informational-masking experiments, however, the target sentence started *after* the competing-speech sentence; this was done to give the listeners some marker as to which sentence to report from the mixture.

II. EXPERIMENT 1

Experiment 1 used the speech-in-speech-in-noise design to measure informational masking. The conditions are summarized in Fig. 1. In the baseline condition ("N"), SRTs for a target sentence (here "Christmas is coming soon") were measured against a masker consisting of a static noise alone (represented by the stippled box). In the experimental condition ("SN"), SRTs were measured against a masker made from both competing speech ("The dirty boy was washing the puppy licked his master") and a static noise. Note that the target sentence started 750 ms after the first of the competing sentences. In a control condition ("MN"), SRTs were measured against a masker made from both a modulated noise (shown by the gray waveform) and the static noise.

Both the modulated noise and the static noise had the same long-term average spectra as the target speech and competing speech. The modulated noise was also designed to be equivalent to the competing speech in terms of its overall fluctuations in power: it was constructed using a one-band vocoder, so lacked the spectral variations in the competing speech and was not recognizable as speech. The level of the competing speech and modulated noise was set to be 4 dB less than the level of the static noise, i.e., a long-term average speech-to-noise ratio of −4 dB. This value was chosen as a compromise between the requirement to minimize the level of the competing speech (so that it added only an insubstantial amount of extra acoustic power to the main noise), and the requirement that the competing speech should be intense enough to be intelligible (in order to generate some informational masking).

Given that the speech and modulated noise were presented at a long-term average speech-to-noise ratio of −4 dB, a simple calculation shows that they would be expected to lead to an increase in power of 1.5 dB over the power of the static noise, on average. The energetic masking in the SN and MN conditions should therefore be 1.5 dB higher than in the N condition. That is, if energetic masking was the only factor affecting SRTs in all three conditions, then the predictions for the SRTs are

$$\mathrm{SRT}_{\mathrm{SN}} = \mathrm{SRT}_{\mathrm{MN}} = \mathrm{SRT}_{\mathrm{N}} + 1.5, \quad (4)$$

where the subscripts refer to the condition. But if the competing speech did indeed add some informational masking (here termed I), then

$$\mathrm{SRT}_{\mathrm{SN}} = \mathrm{SRT}_{\mathrm{MN}} + I = \mathrm{SRT}_{\mathrm{N}} + 1.5 + I. \quad (5)$$

Thus, a comparison of the SRT in the SN and MN conditions enables the amount of informational masking to be calculated.

A. Stimuli

The stimuli were generated from combinations of target speech, static noise, competing speech, and modulated noise. The *target speech* was a sentence taken from the Audiovisual Sentence List (ASL) (MacLeod and Summerfield, 1990) spoken by a male talker. Each sentence had a simple syntactic structure, with three keywords for scoring (e.g., *the mother cooked the dinner*). Following MacLeod and Summerfield (1990) the full set of 270 sentences was divided into sets of 15 sentences. The *competing speech* was a pair of ASL sentences, concatenated without gaps. The pairs of sentences were selected pseudo-randomly so that they would be different from each other and different from the target sentence.

Note that the target speech and competing speech were spoken by the same talker. The *static noise* was a speech-shaped, unmodulated noise. It was constructed by cutting a random section (with 10-ms raised-cosine gates) from a 15-s burst of speech-shaped noise, which was in turn created by an inverse fast Fourier transform of the average spectrum of all the ASL sentences after randomizing the phases. Thus the static noise had the same long-term average spectrum as the target and competing speech. The *modulated noise* was formed from the same speech-shaped noise but was modulated by the envelope of the competing speech (cf. Festen and Plomp, 1990). This envelope was calculated by extracting its instantaneous amplitude (via the Hilbert transform; e.g., Hartmann, 1998) which was low-pass filtered with a cut-off frequency of 32 Hz (with a first-order Butterworth filter with 3-dB/octave slopes). All stimuli were digitally processed at a sample rate of 44.1 kHz.

The maskers were combined into three conditions, termed N, SN, and MN. The N condition consisted of the target speech and static noise alone; in the SN condition, competing speech was added; the MN condition paralleled the SN condition, but modulated noise was added instead of competing speech.

The competing-speech or modulated-noise pairs always started at the same time as the static noise, but the target speech started 750 ms later. The choice of 750 ms was made because it was approximately half the duration of the first of the competing sentences so that the target speech started halfway through the first competing-speech sentence. Thus, in the SN condition, the target speech could be distinguished as the second of three sentences. A short warning tone was presented 750 ms before the onset of the static noise. The static noise was presented at an overall level of 60-dB SPL, and, as noted, the competing speech and the modulated noise were presented 4 dB lower.

After some training, all listeners were able to correctly identify the target speech in competing speech in the absence of static noise, except for three of the most hearing-impaired listeners, who reported that the simultaneous talkers were "jumbled up." This inability to distinguish two talkers at equal levels was interpreted as a raised speech-reception threshold, rather than lack of training, because they were able to follow all other instructions as successfully as the other listeners, and they were all at the higher end of hearing impairment for this study.

Due to the limited availability of suitable experimental spaces, the experiment was run in three different locations: a sound-treated booth and two separate quiet offices. Although these locations had different equipment and different levels of background noise, tests showed that the choice of location did not affect speech recognition in the N condition, so it was assumed that the SN and MN conditions would also be unaffected by the choice of location. Three sets of equipment were used for experimental control and sound generation: in the sound-treated booth, an RME DIGI96/8 PAD soundcard was used with an Arcam A80 amplifier and Sennheiser HD580 Precision headphones, whereas in both offices AudioCapture UA-5 USB audio interfaces were used with EDIROL Audio Capture UA-5 D/A converters, SAMSON C-Que 8 headphone amplifiers, and Sennheiser HD580 Precision headphones. In the sound-treated booth, listeners were able to speak to the experimenter via a microphone; in the offices, the experimenter and listener were in the same room and could speak directly to each other.

B. Procedure

Six-point psychometric functions were measured for speech reception (i.e., the identification of the target speech) as a function of its level relative to that of the masking noise. All stimuli were presented diotically. Listeners were asked to identify the target speech in the stimulus as the second of three sentences heard, and repeat it to the experimenter. The number of keywords accurately repeated was scored, accepting as correct any responses that were homonyms and/or had the same stem as the keyword (MacLeod and Summerfield, 1990). Listeners were encouraged to repeat the sentence they thought they heard, even if it was little more than a guess and even if they thought it made little sense. The levels of the target speech for the six points in each psychometric function were chosen individually on the basis of training data, but always covered a range of 10 dB at 2-dB intervals. The aim was for few words to be correctly identified at the lowest target level, yet for most words to be correctly identified at the highest target level. If necessary, the levels were revised between blocks of trials.

The experiment was run in two sessions, each with nine blocks of trials. A block consisted of 15 trials at each of the six levels. The resulting 90 trials in a block were ordered randomly. The target-speech sentences used in a set of 15 trials were chosen at random, without replacement, from one of the ASL lists of 15 sentences. Within each block, a different list was used for each point of each psychometric function. The type of masker—N, SN, or MN—was fixed within a block but varied across blocks. Three blocks were run for each condition; thus there were 45 trials per point for each listener. This resulted in 135 keywords per point since there are three keywords in each ASL sentence.

A 45-min structured training session preceded data collection in the first session, in which listeners practiced responding to stimuli similar to those used in the experiment. Training began with the N condition, followed by the MN condition, then the SN condition with the competing speech spoken by a female talker, to make it easier to identify the target, and finally the SN condition with both the target and competing speech spoken by the same male talker. At each stage the signal-to-noise ratio of the target was set to a relatively easy level, and then progressively reduced. The training used the Bench–Kowal–Bamford (BKB) sentences (Bench and Bamford, 1979) instead of the ASL sentences for the target and competing speech in order to avoid any memorization of the sentences. The BKB sentences have comparable durations, syntaxes, semantic content, and scoring methods to the ASL sentences.

C. Analysis

An analytic function was fitted to the data using a standard least-squares method to determine quantitatively the

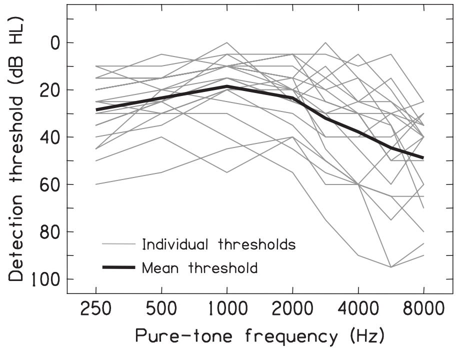

FIG. 2. The pure-tone audiograms of the 20 elderly listeners' better ears in experiment 1 (gray lines) and the corresponding mean audiogram (thick black line).

50% SRT (L_{SRT}, in decibels) and the peak gradient (m, in %/dB) of the psychometric function. The function was a two-parameter cumulative-normal function:

$$W = 100\Phi\left(\frac{\sqrt{2\pi}}{100}m(L - L_{SRT})\right), \qquad (6)$$

where L is the level of the target sentence (decibels), W is the percentage of keywords correctly identified at that level, and $\Phi(z)$ is the cumulative standard normal distribution.[2]

D. Participants

Eight young and 20 elderly listeners took part in the experiment. Seven of the young listeners were students at the University of Strathclyde, Glasgow, and the other was a staff member at the MRC Institute of Hearing Research (IHR). All of the young listeners had self-reported normal hearing. The 20 elderly listeners were volunteers who had previously agreed to take part in research at IHR. They were originally recruited from hearing clinics, the electoral register, and other sources. Figure 2 shows the elderly listeners' audiograms. Their better-ear average ("BEA;" defined as the mean of their audiometric thresholds at 0.5, 1, 2, and 4 kHz) ranged from 6 to 53-dB hearing level (HL), and most of their audiometric functions showed some degree of the high-frequency hearing loss associated with presbycusis. The elderly listeners' BEA was correlated with their age [$r(18) = 0.54$, $p = 0.01$].

E. Results

Figure 3 shows the psychometric functions of each listener for each of the three conditions, with the elderly listeners ordered in terms of their hearing impairment. There was the expected overall effect of hearing loss, in that the psychometric functions of those with greater hearing impairments were generally shifted further to the right than those with lesser hearing impairments. For everyone, the leftmost psychometric function was that of the N condition (circles/solid lines); this was expected as the N condition had the least power of any of the maskers, by 1.5 dB. More importantly, the rightmost psychometric function for *all* listeners was the SN condition, with the MN condition in the middle. That is, there was more total masking in the SN condition than the MN condition. This result is consistent with the prediction that there would be informational masking in the SN condition that is not present in the MN condition.

Figure 4 shows the SRTs [derived from Eq. (6)] for the N condition (top panel), the difference between SRTs in the MN and N conditions (middle panel), and the difference between SRTs in the SN and MN conditions (bottom panel). Note that the scales of the ordinates are the same in all three panels, despite their different ranges. There was a strong positive relationship between SRT_N and hearing impairment [top panel; $r(18) = 0.902$, $p < 0.001$]. But there was no relationship between the $SRT_{MN} - SRT_N$ difference and hearing impairment [middle panel; $r(18) = -0.03$, $p = 0.91$]; across listeners, SRT_{MN} was between 0.9 and 2.8 dB higher than SRT_N, with a mean value of 1.7 dB (SD=0.5 dB). This mean value was close to the predicted value of 1.5 dB given the difference in power of the maskers. Finally, there was a clear negative correlation between $SRT_{SN} - SRT_{MN}$ and hearing impairment [bottom panel: $r(18) = -0.78$, $p < 0.001$]. Across listeners, SRT_{SN} was between 1.0 and 4.6 dB higher than SRT_{MN}, with a mean value of 3.2 dB (SD=0.9 dB). Given the design of the experiment, this $SRT_{SN} - SRT_{MN}$ difference of, on average, 3.2 dB is argued to represent the amount of informational masking due to the competing speech. The negative correlation between it and hearing impairment—which accounted for over half of the variance in the data ($r^2 = 0.61$)—indicates that the listeners with poorer audiometric thresholds showed less informational masking. The effect was relatively small in terms of decibels; however, the values of the $SRT_{SN} - SRT_{MN}$ difference calculated from the regression line at BEAs of 10 and 50 dB were, respectively, 4 and 2 dB.

The effect of age is shown in Fig. 5. The mean $SRT_{MN} - SRT_N$ difference (left pair of bars) was near identical across age group and was statistically insignificant [$t(26) = -0.33$, $p = 0.75$]. The $SRT_{SN} - SRT_{MN}$ difference (right pair of bars) differed by 0.5 dB across age group, with the younger group giving the higher difference. This difference did not quite reach statistical significance across age group [$t(23.52) = 1.97$, $p = 0.06$], although the power of the statistical test was undoubtedly reduced by the relatively large variability for the older group.

Figure 6 shows the individual SRTs from Fig. 4, but plotted against the listeners' ages (the plotted regression lines, and the correlations reported below, are for the older group only). For SRT_N (top panel), there was a trend for older listeners to have higher thresholds than the younger elderly listeners, as would be expected from Fig. 4 given the correlation between age and BEA. However, the scatter in the data was relatively high and the correlation was not significant [$r(18) = 0.38$, $p = 0.10$]. There was, however, an effect of age on the $SRT_{MN} - SRT_N$ difference [top panel; $r(18) = 0.45$, $p = 0.05$], with the oldest listeners showing a greater $SRT_{MN} - SRT_N$ difference, although the differences observed were less than 2 dB. There was no significant effect of age amongst the elderly listeners on the amount of informational masking, $SRT_{SN} - SRT_{MN}$ [bottom panel; $r(18) = -0.37$, $p = 0.11$].

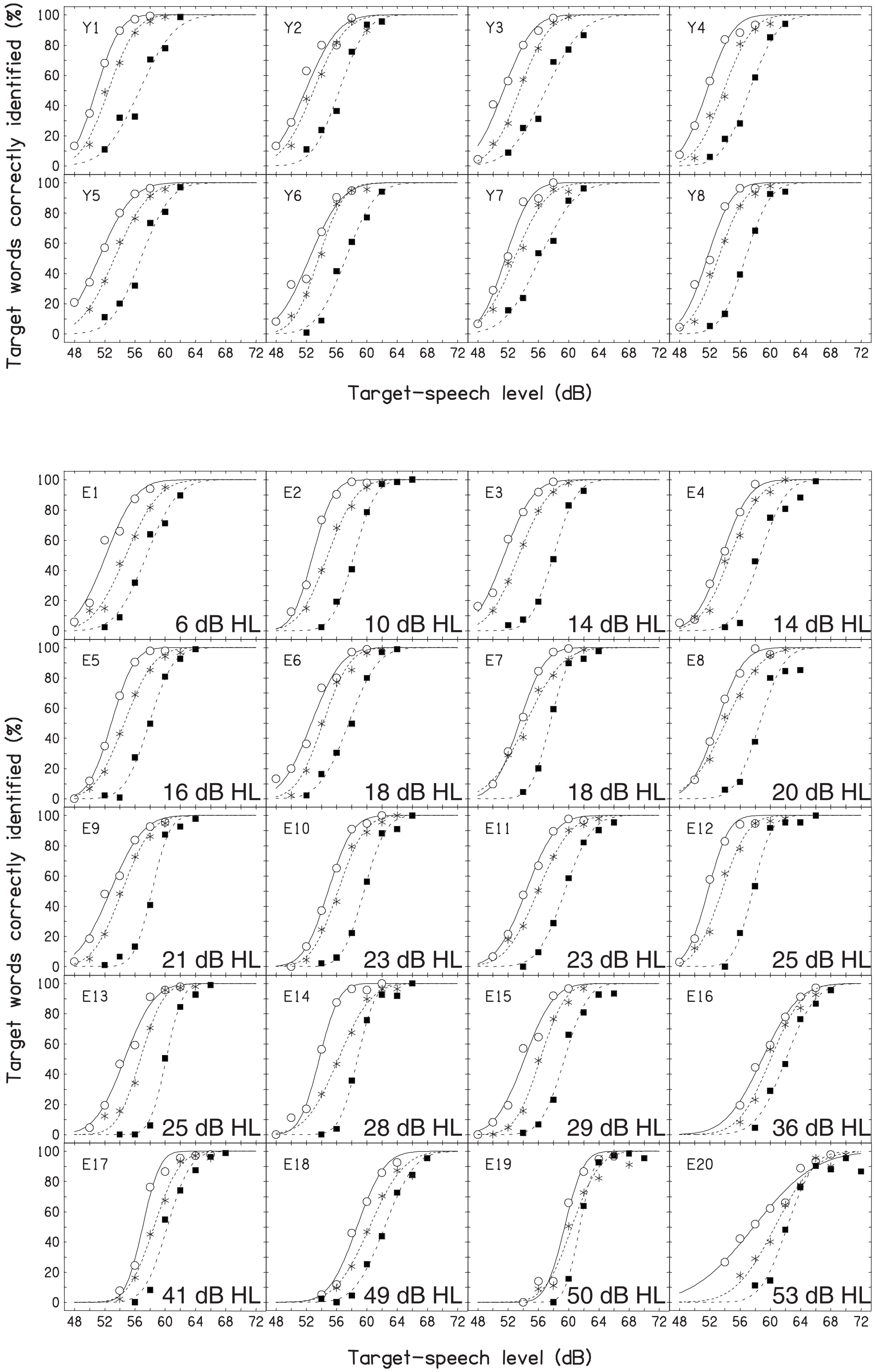

FIG. 3. The psychometric functions measured in experiment 1 for all young listeners (top panels) and elderly listeners (bottom panels). In each panel the symbols show the measured data and the lines show the best-fitting cumulative-normal functions [Eq. (6)]. The parameter is condition: N: open circles and solid lines; MN: asterisks and dashed lines; SN: filled squares and dashed lines. The elderly listeners are ordered by their BEAs, which are shown in the bottom-right corner of each panel.

F. Discussion

Speech-reception thresholds were measured in three conditions of masking: in static noise alone (termed N), in competing speech combined with static noise (SN), and in modulated noise combined with static noise (MN). It was found that the SRTs in the MN condition were, on average,

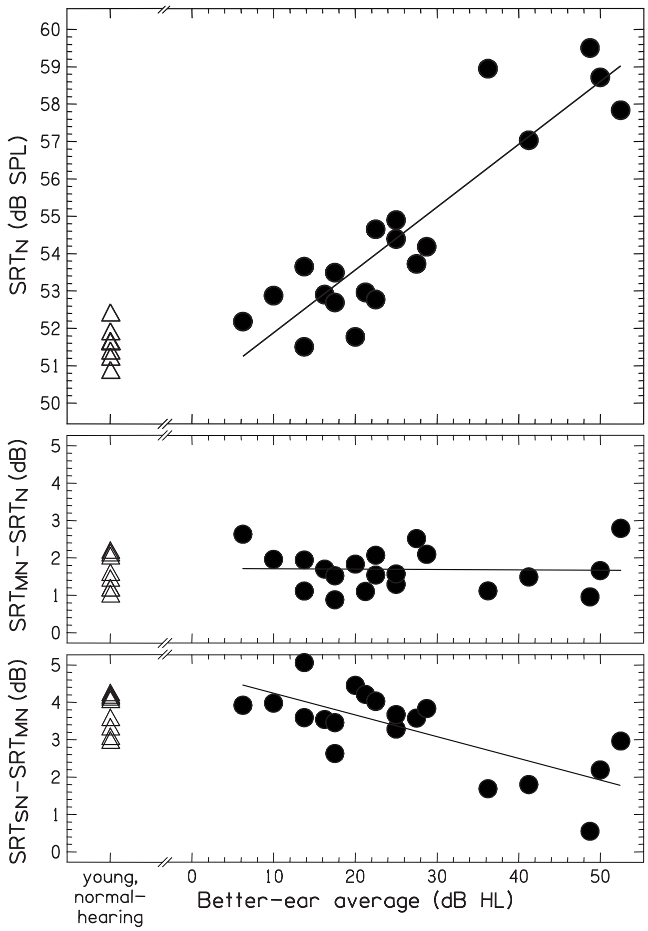

FIG. 4. The derived SRTs and differences in SRTs measured in experiment 1. The top panel shows the speech-in-noise thresholds (SRT$_N$) for each listener; the middle panel, the effect on SRT of adding modulated noise (SRT$_{MN}$−SRT$_N$), and the bottom panel, the amount of informational masking (SRT$_{SN}$−SRT$_{MN}$). The regression lines are fitted through the elderly listeners' data only (filled circles).

1.7 dB greater than in the N condition. This difference can be attributed to energetic masking, as it closely matches the prediction of 1.5 dB made on the basis of the difference in powers of the maskers. It was also found that the SRTs in the SN condition were in turn 3.2 dB greater, on average, than in the MN condition. This difference was attributed to informational masking from the competing speech. Although this was small in terms of decibels, it was large in terms of

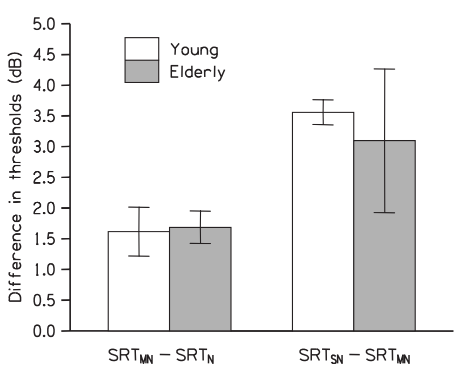

FIG. 5. The across-age-group averages of SRT$_{MN}$−SRT$_N$ and SRT$_{SN}$−SRT$_{MN}$ measured in experiment 1. The error bars are 95%-confidence intervals.

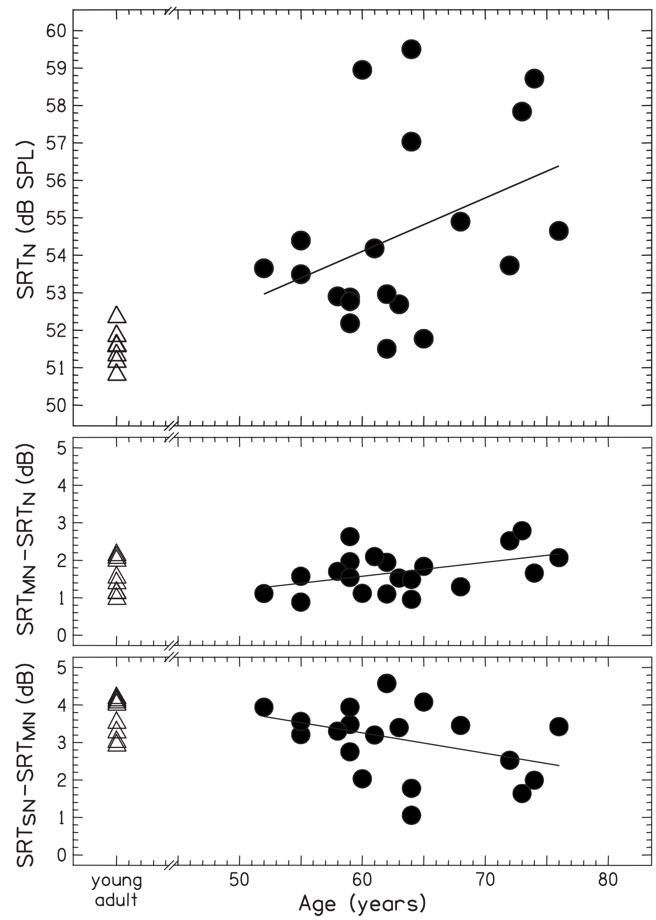

FIG. 6. As Fig. 4, but with the data plotted against age instead of BEA hearing loss.

speech recognition, given that the slope of the psychometric functions was about 10%–20%/dB in this experiment.

An average effect of about 3 dB is at the smaller end of the range of informational-masking effects reported by other researchers (Carhart *et al.*, 1968, 1969a; Freyman *et al.*, 1999; Brungart, 2001; Li *et al.*, 2004; Wu *et al.*, 2004). Of these, Carhart *et al.* (1968) deserves further discussion, as their design was similar to the current experiment although (1) they used a combined masker of speech and modulated noise, not speech and static noise and (2) they used equal-level competing speech and noise masker,[3] instead of reducing the level of the speech in order to minimize its contribution to the energetic masking. They found a difference of 7.8 dB between SRTs for their equivalents of the N and SN conditions. Of this, they attributed 3 dB of the effect to the addition of the competing speech, leaving 4.8 dB for informational masking. This estimate of the amount of informational masking is similar to that observed in the present experiment, despite the differences in sentence materials, choice of noise, and measurement of levels. The issue of the level of the competing speech relative to that of the noise is considered in experiment 2 below.

Attempted quantifications of informational masking have been found to be remarkably dependent on the exact experimental conditions used, and it must be remembered that our results are, strictly, only applicable to the present speech-in-speech-in-noise design. First, the presence of the noise used in the SN condition could itself have affected the amount of informational masking generated by the competing speech: for example, Freyman *et al.* (2001) showed that adding a noise at the same location as the competing talker could reduce its informational masking, and Kidd *et al.* (2005) showed that adding a noise at the same frequencies as a competing talker reduced its informational masking. Second, there could be elements of informational masking induced by the noises themselves: for example, Kwon and Turner (2001) argued that any modulation of maskers could mask speech, which could be considered informational masking, and Lutfi (1990) argued that even masking by an unmodulated noise could include some informational masking due to uncertainty. Third, the amount of informational masking of speech depends on the differences the experimenter allows between the target speech and the competing speech (e.g., Carhart *et al.*, 1969b; Freyman *et al.*, 1999; Brungart, 2001) and may depend on which speech materials are used. It is a wider question as to what laboratory experiments would best represent real-life listening and whether competing speech should be mixed with some other form of masking sound, but in many ways the "worst-case" condition, though experimentally popular, may be unrealistic. It would likely use the same talker for the target and competing speech (or even multiple sources of competing speech; Brungart *et al.*, 2001; Freyman *et al.*, 2001), presented from the same location and at the same level, timed so that they start simultaneously, and would include confusable keywords that are themselves concurrent. In contrast, the antithetical condition to this—with different-gender talkers, from different locations, at different levels, with clear timing differences, and differing syntax and semantics—is likely to give less informational masking, but may also be more representative of real life.

The effects in this experiment were only weakly affected by age. The $SRT_{MN} - SRT_N$ difference did not differ across age group (Fig. 5, left panel), although within the older group it did increase slightly with age, at a rate of 0.37 dB/decade (Fig. 6, middle panel). In contrast, the $SRT_{SN} - SRT_{MN}$ difference was smaller for the older group, and within the older group it decreased at a rate of 0.55 dB/decade. But neither effect reached statistical significance at the usual $p = 0.05$ level, which we attribute to the relatively large scatter in the relevant data points. Although suggestive then that informational masking reduces with age, the results remain somewhat inconclusive.

The effect of hearing impairment was clearer; however, the hearing-impaired elderly listeners were less susceptible to informational masking than the normal-hearing listeners. This result supports that of Arbogast *et al.* (2005), who also observed less informational masking for hearing-impaired listeners. They suggested that this effect may not necessarily be due to a fundamental difference in hearing-impaired listeners' susceptibility, but could instead be related to the design of the experiment. First, hearing-impaired listeners may have been less able to hear the competing speech due to the greater effects of energetic masking, and so it would generate less informational masking. Second, there may be a ceiling effect, such that there is less informational masking when the target speech is at a higher level than the competing speech (see also Brungart, 2001). Both of these applied to our experiment, as (1) we presented the *competing* speech at the same speech-to-noise ratio for both normal-hearing and hearing-impaired listeners, but not the same sensation level, and so the intelligibility of the competing speech may well have been reduced, but (2) we presented the *target* speech at a higher level to the hearing-impaired listeners, as otherwise they would not have reached the 50% criterion for the speech-reception thresholds. Experiment 3 below was, in part, designed to avoid this potential criticism, whereas experiment 2 explored the effect of competing-speech level for normal-hearing listeners.

III. EXPERIMENT 2

Experiment 2 again used the speech-in-speech-in-noise design but measured performance across a range of speech-to-noise ratios (−16 to +4 dB) rather than at just −4 dB. In each of these, the competing speech in the SN condition was predicted to add some informational masking and some additional energetic masking relative to the N condition, but it would be expected that both effects would be minimal when the competing speech was near its own threshold and both effects would be larger when it was far more intense. Also, to check the assumption that the additional energetic masking was the same as the increase in power due to the addition of competing speech, corresponding MN conditions were tested, and the N condition was included as a baseline.

A. Procedure

The stimuli were constructed in the same way as those of experiment 1. In the SN conditions, nine competing-speech levels were used, varied parametrically across blocks at speech-to-noise ratios of −16, −10, −8, −6, −4, −2, 0, +2, and +4 dB. In the MN conditions, five modulated-noise levels were used at modulated-noise-to-static-noise ratios of −10, −4, 0, +4, and +8 dB (these will be referred to below as "noise-to-noise ratios" to simplify the text). Finally, the N condition was the same as in experiment 1, with neither competing speech nor modulated noise. The target speech was again varied parametrically within each block, with target-speech-to-noise levels of −10, −8, −6, −4, −2, and 0 dB.

B. Results and discussion

The procedures and equipment were generally the same as for experiment 1, but rather than repeating the target sentence to the experimenter, the participants typed their responses. The responses were marked by the experimenter off-line after the experiment. There were again three blocks of 90 trials for each condition.

The SN and MN conditions were tested in two separate procedures. For the SN conditions, there were nine competing-speech levels which, with the N condition, resulted in a total of 30 blocks for each listener. The listeners took between four and ten sessions to complete the 30 blocks: the durations of the sessions were based on each listener's preference. Four young listeners participated, aged 24, 24, 32, and 35 years old. Three were students at the

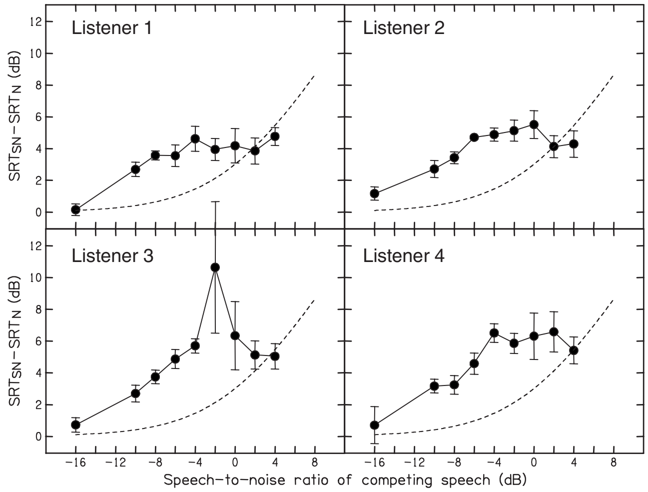

FIG. 7. The SRTs measured in the SN conditions of experiment 2. In each panel the symbols show the measured data and the lines show the predictions given by energetic masking alone, based on the long-term average power of the total masker. The four panels are for the four listeners. The error bars are 95%-confidence intervals.

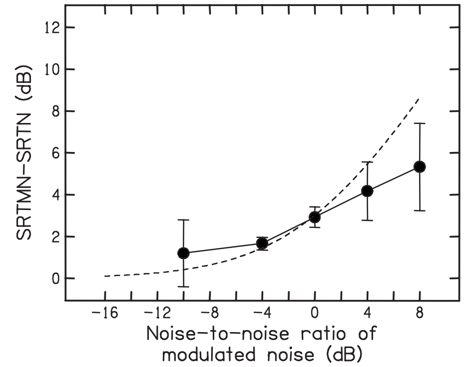

FIG. 8. As Fig. 7, but for the MN conditions of experiment 2. The data are averaged across the listeners. The error bars are 95%-confidence intervals.

University of Strathclyde, Glasgow, and the other was a staff member at the MRC Institute of Hearing Research. For the MN conditions, there were five modulated-noise levels. There were three listeners, who completed the resulting 18 blocks (including the N condition) in three to five sessions each. All three participants were staff members of the MRC Institute of Hearing Research. One also completed the SN conditions of the current experiment, and one was the first author. All listeners in this experiment had self-reported normal hearing.[4]

Figure 7 shows the results for each listener, plotted as the $SRT_{SN}-SRT_{N}$ difference (the values of SRT_{N} were −9.6, −8.5, −8.2, and −10.3 dB for the four listeners). The SRT difference gradually increased as the speech-to-noise ratio was increased, but then remained approximately constant for the highest ratios.[5] The dashed lines show the expected effect if the masking was due *solely* to the total power of the combined masker (e.g., 0.6 dB at −8 dB, 1.5 dB at a speech-to-noise ratio of −4 dB, or 3 dB at a speech-to-noise ratio of 0 dB, and so on). The results—except at the extremes of the range—were generally considerably higher than expected from this model.

Figure 8 shows the corresponding results from the MN conditions, plotted as the $SRT_{MN}-SRT_{N}$ difference. The results are averaged across all three listeners as they responded similarly. For noise-to-noise ratios up to 0 dB, the results were broadly similar to those predicted from the increases in the overall power of the combined masker (dashed line): thus the assumption that the increase in energetic masking was equal to the increase in power was validated. But for higher noise-to-noise ratios, however, the SRTs were *lower* than predicted. We believe that this effect was due to listeners being able to take advantage of the dips in the modulated noise (e.g., Howard-Jones and Rosen, 1993; Cooke, 2006); note that this argument would not be expected to apply at lower noise-to-noise ratios, as there the static noise would have been considerably more effective at filling in the dips of the modulated noise.

The results add to those of experiment 1 in clearly demonstrating that the presence of the competing speech gives additional masking, beyond that which would be expected from its energy, of up to 3–5 dB, depending on the listener. The informational masking at a given speech-to-noise ratio can be estimated as the difference between the values shown in Fig. 7 and those in Fig. 8. The largest amount of informational masking was observed for competing-speech-to-noise ratios in the region of −6 to −4 dB, suggesting that the choice of −4 dB in experiment 1—which was made so that the competing speech would only add a small amount of energetic masking while still being intelligible—was about optimal. Up to this point, the amount of informational masking increased with increasing levels of competing speech, suggesting that in this range the concomitant increase in the intelligibility of the competing speech may directly affect the amount of informational masking obtained. For speech-to-noise ratios greater than −4 dB, however, SRT_{SN} did not increase but SRT_{MN} did. Thus, at higher speech-to-noise ratios, the amount of informational masking (the difference) declined, despite the (presumably) continued intelligibility of the competing speech, to the extent that there was effectively no informational masking at a speech-to-noise ratio of +4 dB. The intelligibility of the competing speech cannot therefore be the *sole* determinant of informational masking.

Rather, the effects are consistent with the notion that informational masking was exacerbated by the similarity of target-speech and competing-speech intensities, suggesting an element of confusion between target speech and competing speech (e.g., Brungart *et al.*, 2001). This has been previously demonstrated by analyzing listeners' errors and counting the words from the competing speech reported instead of those of the target speech: Brungart (2001) and Arbogast *et al.* (2002) found that up to 90% of the incorrect responses involved keywords from the competing speech, although both of their experiments used the coordinate response measure (CRM) speech corpus (Bolia *et al.*, 2000), in which the set of possible responses is extremely limited. In contrast, a corresponding analysis of the present experiment (which used ASL sentences) showed that no more than 4% of the competing-speech keywords were reported at any combination of target-speech and competing-speech levels. Thus our listeners only occasionally confused the competing speech with the target speech to the extent that they would report the wrong keywords. It is likely that part of this reluctance to confuse the target and competing speech is because the present ASL materials can be distinguished, at least in part, by their wide vocabulary and non-identical sentence structures, whereas a simple confusion between target and competing words can play a role in "closed-set" designs such as the CRM task, in which the keywords can be freely confused. If confusion contributes to informational masking for the ASL materials, this must occur at the acoustic or phonemic level rather than at the word level. Whatever the exact reason why the amount of informational masking varies with the level differences between two talkers, the effect leads to potential confounds for many designs when comparing different groups' susceptibility to informational masking. To avoid this, the final experiment used a design in which the target and competing sentences were always presented at the same level.

IV. EXPERIMENT 3

Both experiments 1 and 2 used the speech-in-speech-in-noise design to reduce the energetic contribution of the competing speech while still allowing some informational masking. In this experiment, the energetic contribution from the competing speech was further reduced by separating it in frequency from the target speech, using a method inspired by Arbogast *et al.* (2002). They processed CRM sentences using a 15-band pure-tone vocoder and allocated seven (randomly selected) bands to the target speech and seven others to the competing speech. Arbogast *et al.* (2002) found that the SRT in this condition was, on average, 22 dB greater than the SRT in an acoustically equivalent noise and attributed the difference to informational masking. A similar method was adopted here, except that (1) the stimuli here were filtered by a set of narrow bandpass filters instead of applying a pure-tone vocoder (to help preserve some of the natural speech cues) and (2) similarly filtered masking noises were added as well. The design is summarized in Fig. 9; note that the target speech or competing speech used alternate frequency bands (first and second panels). The masking noise underlying the target speech—termed the "target noise"—was made from the same bands as the target speech (third panel), and likewise for the "competing noise" (bottom panel). These four stimuli were then combined in various ways to make four conditions termed "N," "N+," "SN," and "SN+" (see Table I).

The target speech, target noise, and competing noise were present in all the conditions. In the SN and SN+ conditions, the competing speech was also present (in order to introduce informational masking) and the amount of informational masking was calculated as the difference between the SRTs with and without competing speech. In the N+ and SN+ conditions the competing noise was set to be 14 dB more intense than in the N and SN conditions. This was done in order to estimate any contribution to energetic masking from "cross-excitation" of the other set of frequency bands: because (1) the filtering was done with sharp, but not infinitely sharp, bandpass filters, and (2) listeners' auditory filters are not infinitely sharp either, so some overlap in excitation would be expected between bands. The increase in noise level in the N+ and SN+ conditions meant that the combined power of the competing-speech-plus-noise masker was much larger in the SN+ condition than in the SN condition. That is, if masking was entirely due to the power in the competing channel, without any effect of informational masking, then the SRT would be largest in the SN+ condition.

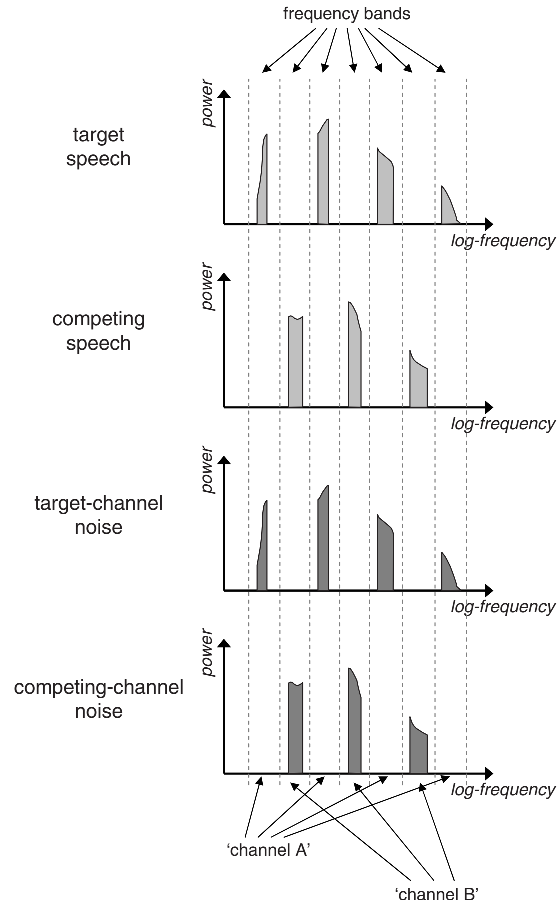

FIG. 9. Illustrative spectra of the stimuli used in experiment 3. A target sentence was filtered into seven frequency bands, of which the first, third, fifth, and seventh were kept and the rest removed (top panel). A competing sentence was similarly filtered but the opposing bands kept (second panel). Two speech-shaped noises were also filtered to form a noise masking the target speech (third panel) and a noise masking the competing speech (fourth panel). All four were then combined into one stimulus for presentation. Not illustrated is a complementary set of stimuli, in which the frequency bands used for the target and competitor were swapped.

TABLE I. The levels of the stimuli used in experiment 3. The levels of the target noise were varied to form the psychometric function. For the young listeners, the levels ranged between 50 and 60 dB SPL, at 2-dB intervals; for the elderly listeners, the levels were chosen individually on the basis of training data, but again always covered a range of 10 dB at 2-dB intervals.

	Levels (dB SPL)			
Label	Target speech	Competing speech	Target noise	Competing noise
N	52	⋯	Varied	52
SN	52	52	Varied	52
N+	52	⋯	Varied	66
SN+	52	52	Varied	66

It was noted earlier that it was likely that the relative intensities of the target speech and competing speech could affect the amount of informational masking. To control for this, the levels of target and competing speech were fixed to be the same throughout the current experiment, and psychometric functions were generated instead by varying the level of the masking noises.

A. Stimuli

The target speech and competing speech were processed versions of the ASL corpus (MacLeod and Summerfield, 1990), selected as for experiment 1; the competing speech was again formed by concatenating two of the ASL sentences, and the target speech started 750 ms after the start of the first competing sentence. The masking noises were always static and initially generated so that they matched the long-term spectra of the speech.

All the stimuli were filtered into seven non-overlapping frequency bands centered on 100, 212, 424, 849, 1697, 3394, and 6788 Hz. Each band was half an octave wide, except for the lowest one, which was widened to 50 Hz (0.74 octaves) to facilitate the design of an appropriate filter. The filters were fourth-order bandpass Butterworth filters with 12-dB/octave slopes, each with less than a 15-ms group delay. Neighboring frequency bands were each separated by half an octave. This separation was designed to minimize the energetic masking of each frequency band by its neighboring bands. Across trials, the bands used for the target speech and competing speech were randomly chosen: either the first, third, fifth, and seventh to the target speech and the second, fourth, and sixth to the competing speech (see Fig. 9) or vice versa (not illustrated). Table I reports the combinations of stimuli used and their levels. Note that these values reported in the table are the long-term average levels of the speech and noise components *before* filtering; the filtering process meant that the total power of what was in the first, third, fifth, and seventh bands was reduced by 5.1 dB, and what was in the second, fourth, and sixth bands was reduced by 9.6 dB. This does not affect the SRTs reported below, which were calculated as the difference in the power of the target speech at threshold and the power of the target noise alone.

B. Procedure

Psychometric functions of target-speech reception against target-noise level were measured. Listeners were asked to listen to the target sentence, which was always the second of three sentences: one of the competing sentences started before it and finished after it. Listeners were required to repeat the target sentence to the experimenter. Six levels of target noise were used to generate each psychometric function, chosen individually for each condition and each listener on the basis of training data. The scoring system, sentence randomization, training, and instructions to the participant paralleled experiment 1.

The experiment was presented in blocks of 90 trials, which lasted 10–15 min. Three such blocks (one per condition in a random order) were presented with short breaks. There were three repetitions of this procedure, spread over 3 days, preceded by a 30-min training session on the first day. Experimenter errors led to three blocks being discarded: for one young listener, the last block was run using the wrong conditions; for two elderly listeners, one psychometric function was measured over a different range to the other two blocks for the same condition. In all three cases, the erroneous blocks were omitted from the analysis, and the estimated performance was based on the average of just two blocks of trials for the three listeners in the affected conditions.

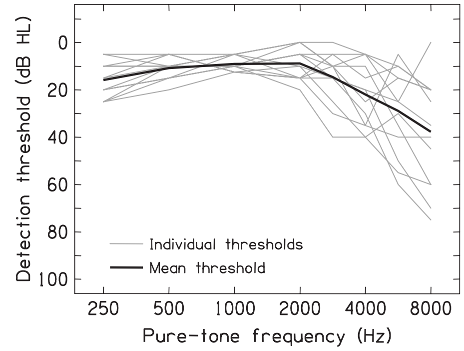

FIG. 10. The pure-tone audiograms of the 20 elderly listeners' better ears in experiment 1 (gray lines) and the corresponding mean audiogram (thick black line).

The stimuli were presented through a 24-bit D/A converter (an RME DIGI96/8 PAD soundcard) and an Arcam A80 amplifier. Participants listened through Sennheiser HD580 Precision headphones while in a sound-treated booth. Participants' responses were picked up by a microphone in the booth and presented over headphones to the experimenter outside the booth. The experimenter recorded the total number of target keywords correctly identified.

C. Participants

Eight young listeners and 13 elderly listeners participated in the experiment. Three of the young listeners were staff at the MRC Institute of Hearing Research and had previously taken part in experiments using the same speech corpus. The other five were recruited from the University of Strathclyde Psychology Department and had no previous experience with psychoacoustical experiments. The 13 elderly listeners were initially invited from population surveys based on the electoral register. They had participated in previous auditory experiments, but these experiments did not involve the ASL corpus. Figure 10 shows the elderly listeners' better-ear audiograms. Their mean age was 66 years (ranging from 51 to 80 years), and their mean BEA was 13-dB HL (ranging from 4- to 21-dB HL); thus most, if not all, would be classified as having normal hearing, though with some high-frequency hearing loss.

D. Results

Figure 11 shows the psychometric functions for six illustrative listeners. The four panels per listener are for the four conditions, and within each panel the two psychometric functions are for the case where the target speech used the even-numbered frequency bands (solid symbols) or where it

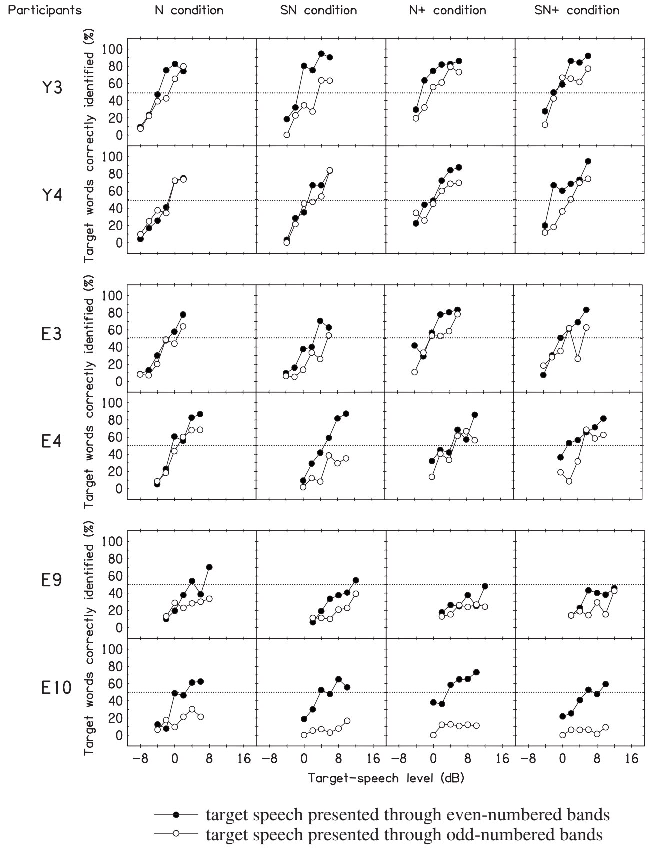

FIG. 11. The psychometric functions from experiment 3 for an illustrative selection of listeners (see text for choices). The conditions are ordered by column. In each panel one psychometric function is for when the target speech used the first, third, fifth, and seventh frequency bands, the other for when it used the second, fourth, and sixth bands.

used the odd-numbered bands (open symbols). For the young listeners (e.g., listeners Y3 and Y4 in Fig. 11), most of the psychometric functions were similar for the two target channels, and a good level of speech understanding was achieved at the higher speech-to-noise ratios. The same occurred in most of the elderly listeners (e.g., E3 and E4 in Fig. 11), but some others had relatively low scores for target speech in one or both of the target channels, not always reaching the 50% criterion for the speech-reception threshold (e.g., E9 and E10 in Fig. 11).

Figure 12 shows the SRTs derived from the psychometric functions via Eq. (6), plotted against hearing impairment (left panels) or age (right panels). The lines are regression lines and are calculated within the elderly group only. Note that some of the SRTs are extrapolations as performance never reached 50%. In addition, any extrapolated SRTs that were improbably large (i.e., derived from particularly shallow psychometric functions) have been capped at +12 dB. The mean SRTs for the younger group of listeners were −3.3, 1.0, −2.0, and −1.4 dB for the N, SN, N+, and SN+ conditions, respectively, and a repeated-measures ANOVA confirmed the overall effect of condition for these younger listeners $[F(3,21)=20.32,\ p<0.001]$. A comparison of the SN and N conditions showed that there was about 4 dB of informational masking ($SRT_{SN}-SRT_{N}=4.3$ dB). Moreover, the SN+ condition did not give the highest SRT despite having the highest overall power. In fact, raising the level of the competing noise in the presence of the competing speech actually reduced the SRT (i.e., $SRT_{SN+}-SRT_{SN}=-2.4$ dB), although on its own the change in level of the competing masker led to an increase in SRT of about 1 dB (i.e., $SRT_{N+}-SRT_{N}=1.3$ dB).

The calculations for the corresponding mean SRTs for the elderly listeners excluded four listeners who gave at least one capped SRT: without them, the means were 1.2, 6.2, 4.2, and 4.8 dB for the N, SN, N+, and SN+ conditions, respectively. Thus, the effect of informational masking ($SRT_{SN}-SRT_{N}$) was 5.0 dB, the effect of raising the competing noise on informational masking was −1.4 dB ($SRT_{SN+}-SRT_{SN}$), and the effect of increasing competing noise in the absence of competing speech was 3 dB ($SRT_{N+}-SRT_{N}$). All three differences were about 1–2 dB larger than the corresponding differences in the younger group. A mixed ANOVA on SRT_{SN}, SRT_{N+}, and SRT_{SN+}, each relative to SRT_{N}, confirmed an overall effect of age group $[F(1,15)=5.44,\ p<0.03]$ and an effect of condition $[F(2,30)=18.01,\ p<0.001]$. The effect of age is supported by inspection of the regression lines within the elderly group, all of which had a positive gradient. The ANOVA did not show an interaction between age group and condition $[F(2,30)=0.83,\ p=0.45]$. However, the amount of informational masking was not itself significantly different between the two age groups $[SRT_{SN}-SRT_{N},\ t(15)=1.34,\ p=0.20]$. The effect of increasing the level of the competing-channel noise was also insignificant across age group $[SRT_{N+}-SRT_{N},\ t(15)=1.83,\ p=0.09]$, but when combined with competing speech it was significant $[SRT_{SN+}-SRT_{N},\ t(15)=2.17,\ p=0.05]$. As noted, four of the elderly listeners were excluded from these calculations, as they gave at least one capped SRT. It is intriguing that these people were primarily the oldest and more impaired of the group: they were the oldest four, included four of the five highest hearing levels, and returned three of the four highest values of SRT_{N}.

E. Discussion

The difference in SRTs between the N and SN conditions was about 4–5 dB. Given that the only experimental difference between the two conditions was that there was competing speech present in the SN condition but not in the N condition, and assuming that its presence did not affect the energetic masking of the target speech because of the separation of frequency bands, then this difference in SRT can be attributed to informational masking.

We did not find any evidence that, on average, the amount of informational masking increased with age. Nevertheless, it will be noted that it was the oldest listeners who gave high SRTs that required capping, and as some of the capped SRTs were in the SN condition but none were in the N condition, then these listeners may well have been particularly susceptible to informational masking.

The amount of informational masking observed here was about 1–2 dB larger than that measured in experiment 1. This may have been because the target speech and competing speech were presented at the same intensity (the pre-

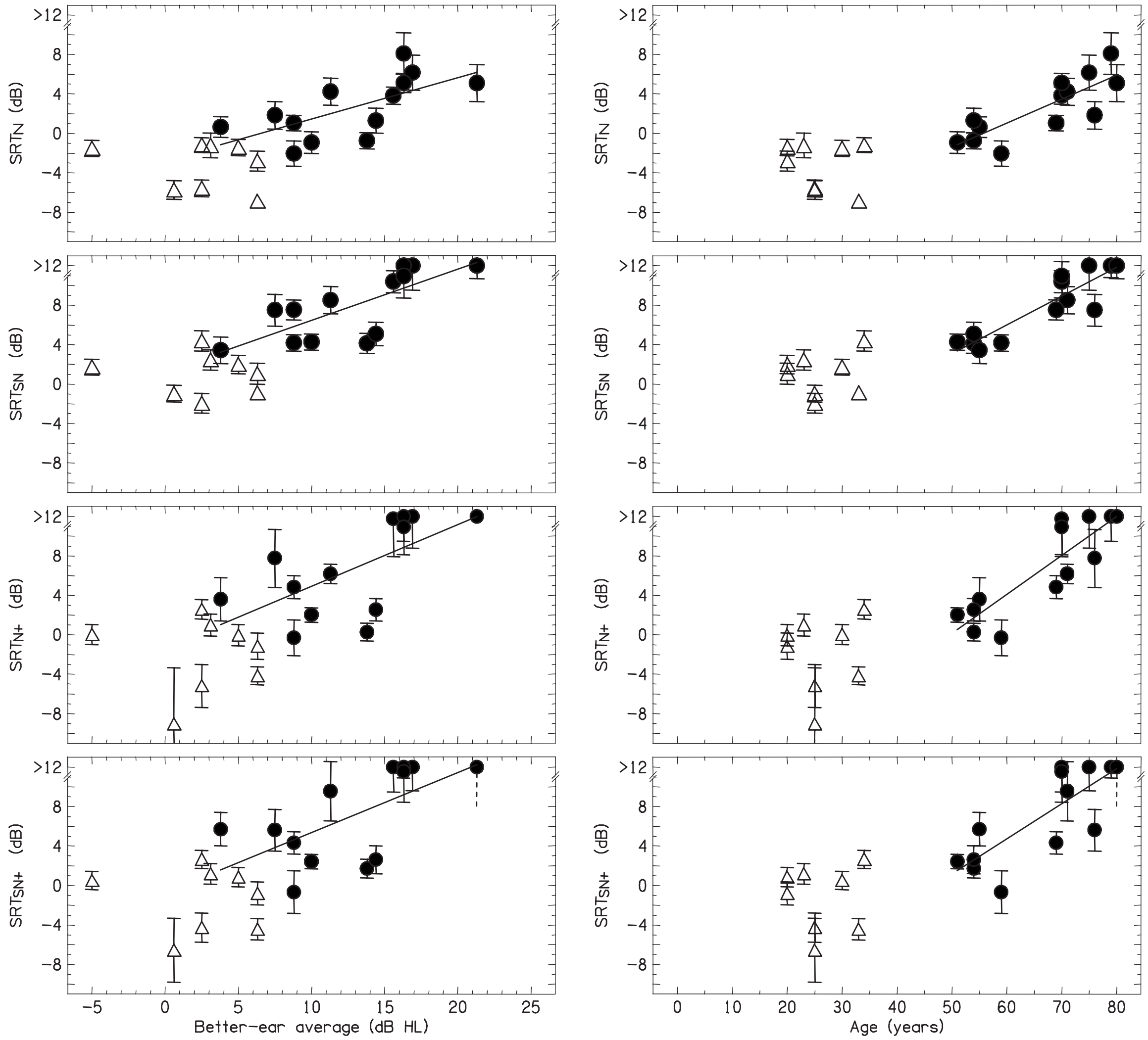

FIG. 12. The SRTs measured in experiment 3. The conditions are ordered by row. The data are plotted as a function of hearing level (left panels) or age (right panels; any SRT larger than 12 dB was capped at that value (see text). The data from the young listeners are shown as open triangles; the data from the elderly listeners as filled circles. The regression line was fitted to elderly listeners' data only (solid line). The error bars show 95%-confidence intervals; a dashed line indicates an error bar that was too large to show on the graph. Note that the range of BEAs is considerably smaller than that shown in Fig. 4.

vious experiment suggested that differences in intensity could serve as a potential cue to release informational masking). The speech-to-noise ratio in the competing channel was also fixed, but the speech-to-noise ratio in the target channel was varied. It seems unlikely that these differences in signal-to-noise ratio provided a cue with which to distinguish the target and competing speech, as informational masking was spread relatively widely across the range of speech-to-noise ratios tested: there were no obvious dips in the speech-recognition psychometric functions at 0 dB (cf. Egan *et al.*, 1954) or plateaus below 0 dB (cf. Brungart, 2001). It could be that much of the informational masking had already been released by other cues, such as the spectral differences between the target and competing speech that were introduced by the filtering processes, although that would suggest that the amount of informational masking occurring with unfiltered speech could be significantly greater than the 4–5 dB observed here or the 3–4 dB observed in experiment 1.

The amount of informational masking observed here was far less than the 22 dB observed by Arbogast *et al.* (2002), whose method we adapted. Nevertheless, the differences between the two experiments were substantial: (1) Arbogast *et al.* (2002) used pure-tone vocoded speech, not narrow-band filtered speech; (2) their vocoded speech was formed from many combinations of carrier frequencies, whereas our filtered speech was always formed from one of just two sets of frequency bands; (3) they estimated informational masking as the difference in SRTs between (in the present terminology) an N condition and an S condition (not an SN condition); and (4) they used the closed-set CRM corpus, not the open-set ASL corpus. Of these, we believe that the primary determinant of their 22-dB effect was their use of the CRM corpus, as large amounts of informational masking have also been observed in other experiments using this corpus (e.g., 12 dB by Brungart, 2001). Furthermore, in subsequent tests we conducted, using CRM sentences that were band-filtered in a similar fashion to the current experiment, we observed effects of up to 30 dB for normal-hearing listeners between the N and S conditions (Agus, 2008). These comparisons highlight the extent to which methodological differences affect measures of informational masking.

Our calculation of the amount of informational masking here is only strictly valid if the competing speech and target speech were ideally isolated from one another; that is, that they did not overlap in excitation. The comparison of SRT_{N+} and SRT_N demonstrated that increasing the level of the competing noise by 14 dB led to an increase in masking of 1–3 dB (SRT_{N+} versus SRT_N). This result suggests that there was indeed some overlap in excitation between the frequency bands: if there was none, then the $SRT_{N+}-SRT_N$ difference would have been zero. But if a 14-dB change in the level of the noise led to only about a 1–3-dB change in SRT, then the 3-dB increase in overall power due to adding the competing speech would be expected to have only a minimal effect on SRT (the effect was quantified by Agus, 2008 and found to be just 0.1 dB). The overlap in excitation can therefore be neglected in deriving the amount of informational masking. But the addition of the 14-dB change in level of the competing noise led to a *reduction* in informational masking of about 1–2 dB (i.e., SRT_{SN+} versus SRT_{SN}). This suggests that there was less informational masking in the SN+ condition than the SN condition, as the additional competing noise would only have increased the energetic masking of the target speech. This result parallels those of Kidd *et al.* (2005) and Freyman *et al.* (2001), who also found that the addition of a noise reduced the amount of informational masking induced by competing speech, presumably by energetically masking the competing speech. These results highlight an important contrast between the two kinds of masking: an increase in the power of a noise always raises energetic masking but can reduce informational masking.

V. SUMMARY

Three experiments were conducted to study the informational masking for target speech by competing speech. All used a speech-in-speech-in-noise method, designed to conceptually separate the energetic-masking and informational-masking contributions to the speech-reception threshold for target speech. The stimuli were sentences chosen from the ASL corpus. Across all experiments, the amount of informational masking observed was of the order of 3–5 dB. Both experiments 1 and 2 used unprocessed sentences for both the target and competing speech: Experiment 1 demonstrated that elderly listeners were no more susceptible to informational masking than younger listeners, but hearing-impaired listeners showed typically 1–2 dB less informational masking, while experiment 2 showed that the amount of informational masking observed was largest when the levels of the target speech and competing speech were about equal. Experiment 3 used sentences that were narrow-band filtered into separated, non-overlapping frequency bands. Again there was no significant effect of age group on informational masking, although there were some indications that the oldest listeners may have been the most susceptible to informational masking. The present results are consistent with previous experiments (Tillman *et al.*, 1973; Li *et al.*, 2004; Arbogast *et al.*, 2005), showing that any detrimental effects of age and hearing impairment on informational masking of speech by speech are, at most, small.

ACKNOWLEDGMENTS

T.R.A. was funded by a Ph.D. studentship from the Medical Research Council. The Scottish Section of IHR is supported by intramural funding from the Medical Research Council and the Chief Scientist Office of the Scottish Government. We thank Richard Freyman (Associate Editor) and two anonymous reviewers for their insightful comments on the submitted manuscript.

[1]Note that the addition of the static noise also reduces the scale of the fluctuations in power compared to speech alone, and so reduces the opportunity for dip-listening (e.g., Howard-Jones and Rosen, 1993; Cooke, 2006).

[2]This equation is the same as Qin and Oxenham (2003), except that it has been recast so that the slope parameter represents the peak gradient in %/dB.

[3]Carhart used the "frequent peaks" of a VU meter, not a long-term rms average, to compare the levels of speech to noise.

[4]Practical constraints prevented us from running this experiment on hearing-impaired or elderly listeners, but a comparison of similar data for different listeners would clearly be of interest.

[5]There was one outlier, namely, listener 3's higher SRT at the speech-to-speech ratio of −2 dB. In the underlying psychometric function for this speech-to-speech ratio, the listener never scored above 40%. Thus the SRT, estimated at 50% speech reception, was an unreliable extrapolation, leading to a large 95%-confidence interval.

Agus, T. R. (**2008**). "Informational masking of speech for elderly listeners," Ph.D. thesis, University of Strathclyde, Glasgow, UK.

Agus, T. R., Akeroyd, M. A., Nobel, W., and Bhullar, N. (**2009**). "An analysis of the masking of speech by competing speech using self-report data," J. Acoust. Soc. Am. **125**, 23–26.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969a**). "Perceptual masking in multiple background sounds," J. Acoust. Soc. Am. **45**, 694–703.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969b**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411–418.

Carhart, R., Tillman, T. W., and Johnson, K. R. (**1968**). "Effects of interaural time delays on masking by two competing signals," J. Acoust. Soc. Am. **43**, 1223–1230.

CHABA (**1988**). "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859–895.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Durlach, N. I. (**2006**). "Auditory masking: Need for improved conceptual structure (L)," J. Acoust. Soc. Am. **120**, 1787–1790.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**, 2984–2987.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). "Some factors affecting multi-channel listening," J. Acoust. Soc. Am. **26**, 774–782.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial and qualities of hearing scale (SSQ)," Int. J. Audiol. **43**, 85–99.
Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer-Verlag, New York).
Hawkins, J. E., Jr., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.
Hornsby, B. W. Y., Ricketts, T. A., and Johnson, E. E. (**2006**). "The effects of speech and speech-like maskers on unaided and aided speech recognition in persons with hearing loss," J. Am. Acad. Audiol **17**, 432–447.
Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**, 2915–2922.
Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," J. Acoust. Soc. Am. **120**, 2926–2937.
Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.
Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.
Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (**2004**). "Does the informational content of an irrelevant source differentially affect spoken word recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091.
Lutfi, R. A. (**1990**). "How much informational masking is informational masking?," J. Acoust. Soc. Am. **88**, 2607–2610.
MacLeod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29–43.
Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Tillman, T. W., Carhart, R., and Nicholls, S. (**1973**). "Release from multiple maskers in elderly persons," J. Speech Hear. Res. **16**, 152–160.
Tun, P. A., O'Kane, G., and Wingfield, A. (**2002**). "Distraction by competing speech in younger and older listeners," Psychol. Aging **17**, 453–467.
Tun, P. A., and Wingfield, A. (**1999**). "One voice too many: Adult age differences in language processing with different types of distracting sounds," J. Gerontol. B Psychol. Sci. Soc. Sci. **54B**, 317–327.
Watson, C. S. (**2005**). "Some comments on informational masking," Acta Acust. **91**, 502–512.
Wu, X., Wang, C., Chen, J., Hongwei, Q., Li, W., Wu, Y., Schneider, B. A., and Li, L. (**2005**). "The effect of perceived spatial separation on informational masking of Chinese speech," Hear. Res. **199**, 1–10.