Exploring the factors predictive of informational masking in a speech recognition task

Frederick J. Gallun, and Anna C. Diedesch

Citation: Proc. Mtgs. Acoust. **19**, 060145 (2013); doi: 10.1121/1.4799107
View online: https://doi.org/10.1121/1.4799107
View Table of Contents: https://asa.scitation.org/toc/pma/19/1
Published by the Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 3aSCb: Components of Informational Masking

3aSCb3. Exploring the factors predictive of informational masking in a speech recognition task

Frederick J. Gallun* and Anna C. Diedesch

***Corresponding author's address: National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Rd. (NCRAR), Portland, OR 97239, Frederick.Gallun@va.gov**

The effects of informational masking (IM) can be recast as a question of which cues to sound source identity (auditory object formation) are most useful for overcoming IM. We hypothesize that individual differences are related to specific interactions of stimulus and listener-specific variables that determine the effectiveness of the auditory object formation process. Results from our laboratory generally support the well-established relationship between performance and stimulus variables such as spectrotemporal cues (in this case, voice differences) and spatial cues (talker locations). In addition, the listener-specific variables of age and hearing loss were found to interact with the stimulus variables and to be correlated with potential mediating variables such as interaural time sensitivity and minimum levels at which speech identification was possible. Future work will involve developing predictive models that focus on identifying the mediating variables responsible for increased susceptibility to IM and efficient tests to reveal these relationships in individual listeners. The clinical relevance of the ability to identify factors predictive of IM susceptibility will be discussed, including the potential for improved fitting of hearing aids and cochlear implants.

Published by the Acoustical Society of America through the American Institute of Physics

Received 17 Jan 2013; published 2 Jun 2013

INTRODUCTION

Despite decades of research on speech understanding in noise, there is no clinical or theoretical explanation for the increased difficulties some listeners have in multitalker environments as compared to other listeners with similar audiograms and of similar age. This fundamental gap in both our theoretical and clinical knowledge makes it impossible to predict a listener's ability to communicate in restaurants, the workplace, and other social environments with the current clinical tests available. The effects of such communication difficulties on human health include increased self-perception of handicap and social withdrawal (Noble and Gatehouse, 2006).

One of the major obstacles to effective clinical diagnosis is that the current models of the effects of noise on the auditory system are based on "energetic masking" (EM), but the situations described above are dominated by multiple sound sources and should be characterized as "informational masking" (IM; reviewed in Kidd et al., 2008). EM is usually taken to mean that a reduction in performance can be predicted with energy-based models of the peripheral auditory system such as that proposed by Green and Swets (1974, Chapter 8). Such a model usually consists of a bank of filters followed by rectifiers, an integrator stage, and a decision stage based on differences in energy at the output of the integrator. IM is generally invoked in cases where such a model fails to accurately predict the data and no higher process (such as a binaural comparison or pitch extraction stage) can salvage the prediction by being added to the model.

It is reasonable to suggest that while many of the impacts of hearing loss can be captured by known sensory impairments applied to an energy detector model, the current difficulties in explaining individual differences of listeners to hear successfully in noisy environments should be recast as a question of which cues to sound source identity (auditory object formation) are most useful for overcoming IM, and how aging and hearing loss impact the ability to detect and utilize these cues. It is hypothesized that two fundamental differences among listeners are the ability to benefit from spatial separation of sound sources ("spatial cues") and the ability to benefit from differences in the spectral and temporal characteristics of the sound sources ("spectrotemporal cues").

The experiments described here represent an attempt to begin to answer some of our many questions regarding the impacts of age and hearing loss on the ability to benefit from spatial and spectrotemporal cues.

EXPERIMENT 1

A set of data was obtained in which listeners who varied in age and hearing sensitivity were tested on the ability to understand one target talker in the presence of two competing masking talkers. Participants identified keywords associated with the target while attempting to ignore two maskers presented either from the same location as the target or from spatially separated locations. The corpus was the Coordinate Response Measure (CRM) developed by Bolia et al. (2001). Each sentence in the corpus contains one of eight callsigns (Ringo, Charlie, Baron, Arrow, Hopper, Laker, Eagle, Tiger), a color word (Red, White, Green, Blue) and a number word (1, 2, 3, 4, 5, 6, 7, 8), all embedded in the carrier phrase "Ready [callsign] go to [color] [number] now." All CRM sentences were recorded by four males and four females speaking naturally.

Data were collected in an anechoic chamber outfitted with an array of 24 loudspeakers equally separated by 15 degrees. Speech recognition was measured for a target sentence presented among two masking sentences. Target talkers could be male or female, as could masking talkers. All three sentences were drawn from recordings contained in the CRM corpus. The target was presented from the center loudspeaker and the competing speech was presented either from the same loudspeaker as the target (referred to as the 0 degree separation or 'co-located') or from two loudspeakers placed 15, 30, or 45 degrees to the left and right of the center loudspeaker ('spatially separated' conditions). Performance was assessed by fixing the RMS level of the target at 50 dB SPL and adaptively tracking the combined masker level that resulted in 50% correct recognition. The ability to recognize a CRM sentence in quiet was measured as a function of RMS level using the same adaptive tracking procedure but changing the level of the target with no maskers present. The level associated with 50% correct recognition was used to define the Speech Recognition Threshold (SRT) in dB SPL.

Participants were thirty-four Veterans and non-Veterans varying in age (25-74 years) and pure-tone sensitivity (averages for .5, 1, and 2 kHz [PTA_{LOW}] of 4-32 dB HL and averages for 1, 2, and 4 kHz [PTA_{HIGH}] of 3-48 dB HL). Age and PTA_{HIGH} were significantly correlated (r^2=.20), while age and PTA_{LOW} were not significantly correlated (r^2=.04). The seven listeners aged 25-49 with audiometric thresholds of 20 dB HL or better at all measured frequencies were designated as the control group, and performance of the remaining listeners were compared to the mean and standard deviation of performance for the control group.

Although age was correlated with hearing loss at 4-8 kHz in the sample tested, the ability to correctly report the keywords of the CRM stimuli in a quiet anechoic chamber (SRT) was not significantly correlated with age ($r^2 = .04$). This suggests that, due to the use of a closed set of very familiar words, thresholds in the frequencies below 4 kHz, which were also uncorrelated with age, may be more important than are higher frequencies for establishing the basic intelligibility of the CRM stimuli.

Target-to-masker ratios, averaged across all listeners, are shown as a function of target gender, masker gender, and spatial separation in Figure 1. A repeated-measures analysis of variance was conducted, with target-to-masker ratio (in dB) as the dependent variable. The factors of spatial separation (0, 15, 30, or 45 degrees), target gender (male or female), and masker gender (same or different from target gender) were all statistically significant ($p < .001$), but none of the interactions were significant ($p > .05$). Age and SRT were entered as covariates, and both were statistically significant ($p < .001$). Two significant three-way interactions were observed. The first was among age, masker type, and spatial separation ($p < .005$), and the second was among SRT, masker type, and spatial separation ($p < .05$).

Data for the co-located (0 degrees) and spatially separated (45 degrees) conditions for male targets and male maskers are presented as a function of SRT in the top two panels of Figure 2. Older listeners (aged 50-74 years) are shown as square symbols and younger listeners (aged 25-49 years) are shown as triangles. In the bottom panels, the same relationships are plotted for the conditions in which the maskers were female.

In order to better understand the effects of age and hearing loss (operationally defined as the SRT), the relationships shown in Figure 2 were examined using stepwise linear regression. Correlations with each factor alone and with the model based on a linear combination of the two factors are shown in Table 1.

The spatially separated data in Figure 2 can be compared to those of Marrone et al. (2008) as well as to those of Neher et al. (2011), both of which used these same spatial configurations and presented sentences composed of a callsign followed by keywords. The differences in performance between the co-located and spatially separated conditions in both studies were similar to those observed here, despite methodological differences that most likely account for variations in the absolute thresholds, especially as compared with Neher et al. (2011). The range of spatial benefit for the hearing impaired listeners in Figure 2 is similar to that reported in those two studies, although the number of listeners achieving very small amounts of benefit is lower in these data, probably due to fewer examples of moderate to severe hearing loss. The data suggest that neither age nor hearing loss capture the full variability in performance in either of the CRM tasks.

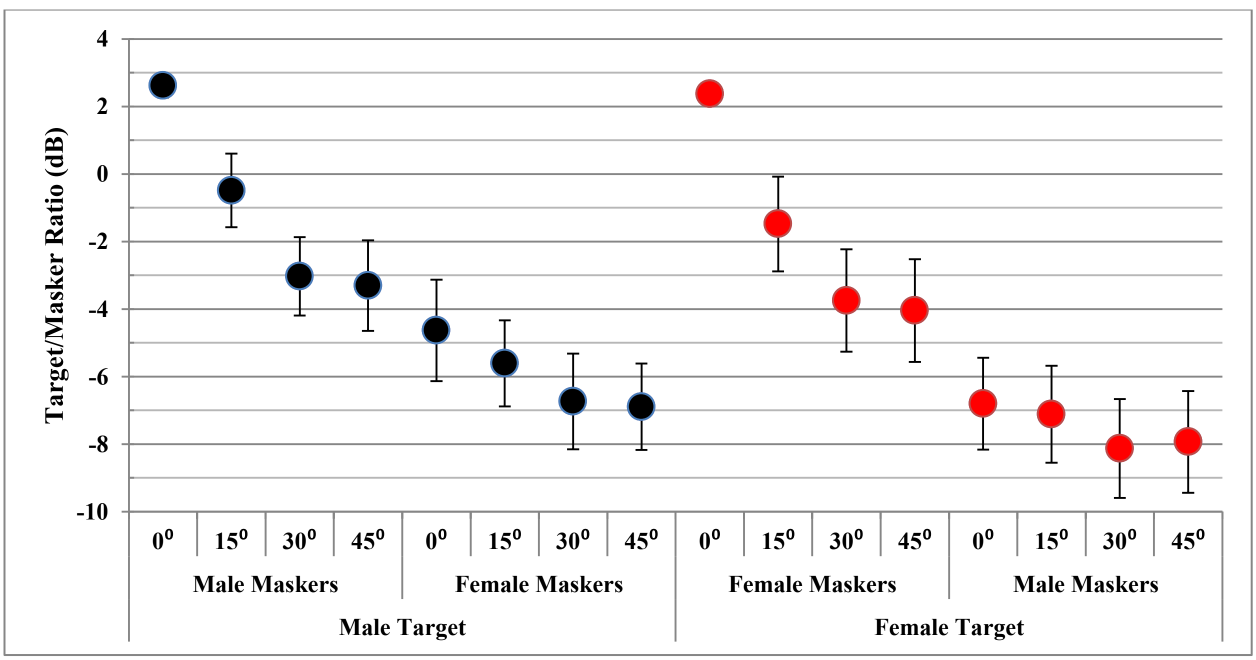

FIGURE 1. Target/masker ratios at threshold for 34 listeners varying in age and hearing loss across 16 combinations of target gender, masker gender, and spatial separation. Error bars indicate ± 2 standard errors of the mean.

TABLE 1. Correlation values associated with the individual predictors (SRT and Age) as well as the correlation of the multiple-factor linear regression model including both SRT and Age (SRT+Age). Values in bold represent the greatest correlation obtained across the three approaches. Correlations marked with a single asterisk (*) are statistically significant at level of p<.05. Correlations marked with two asterisks (**) are significant at a level of p < .001. All significance values are based on a sample size of n=34.

Target	*Male*	*Male*	*Male*	*Male*	*Female*	*Female*	*Female*	*Female*
Masker	*Male*	*Male*	*Female*	*Female*	*Female*	*Female*	*Male*	*Male*
Spatial Separation	*0°*	*45°*	*0°*	*45°*	*0°*	*45°*	*0°*	*45°*
SRT	0.321*	0.677**	0.324*	0.621**	0.540**	0.569**	0.558**	0.595**
AGE	**0.524****	0.431*	**0.467****	0.415*	0.453*	0.504**	0.549**	0.497*
SRT+AGE	0.524**	**0.736****	0.467**	**0.682****	**0.639****	**0.688****	**0.708****	**0.702****

EXPERIMENT 2

The interaural difference in time (ITD) is an acoustical property of the relationship between the spatial positions of a listener's ears and the location of a sound source. When the ears are equally distant from the sound source, the difference in the time of arrival of a sound across the ears is zero. When a sound source is maximally displaced to the left or right of the listener, the ITD can take on values as great as 700 μs, where the maximum interaural difference is determined by the size of the listener's head.

In this experiment, thresholds were obtained by adaptively varying the ITD applied to a train of 4 Gaussian-windowed 2000 Hz tones, each 4 ms in duration. Listeners heard a stimulus with a 0 μs ITD followed by a comparison stimulus with an ITD between 10 and 1500 μs favoring either the right or left ear. When the ITD favored the right ear, the listener would be correct if their response was "right" and incorrect if their response was "left". Using Levitt's (1971) 2-down/1-up formulation, the ITD was adaptively varied and a threshold was calculated.

Thirty-eight listeners participated, varying in age (21-74 years) and hearing thresholds for a 2000 Hz pure tone (0-45 dB HL). Although age and pure-tone thresholds at 2000 Hz were uncorrelated (r^2=.001), as were hearing threshold and ITD at threshold (r^2=.04), a significant correlation was found between age and ITD (r^2=.25). These data suggest that while age alone can reduce sensitivity to ITD, some older listeners do have binaural sensitivity similar to their younger counterparts. This is also demonstrated in Figure 3, where the relationship between ITD and speech recognition thresholds for the male targets and male maskers with a 45 degree separation are plotted for the 17 listeners who also participated in Experiment 1.

Correlations between the ITD thresholds (log transformed) and target/masker ratio thresholds for the two combinations of target and masker type crossed with two of the spatial configurations (0 and 45 degrees) are shown in Table 2. Significant correlations were obtained for both of the same-gender maskers at 45 degrees (p < .05) and for the female targets with female maskers at 0 degrees (p < .05). Correlations between ITD and target/masker ratios for the other target and masker combinations and spatial configurations were not statistically significant (p > .05).

These data suggest that hearing loss and spatial sensitivity are both potential influences on performance, with age potentially acting as a mediating variable by reducing sensitivity to spatial cues. A more detailed description of these relationships requires the use of a wider range of tests and a larger sample with greater variability in age and hearing sensitivity, but these data support the view that the ability to comprehend speech in a complex acoustical environment with multiple talkers is strongly influenced by individual listener characteristics beyond those captured by the audiogram.

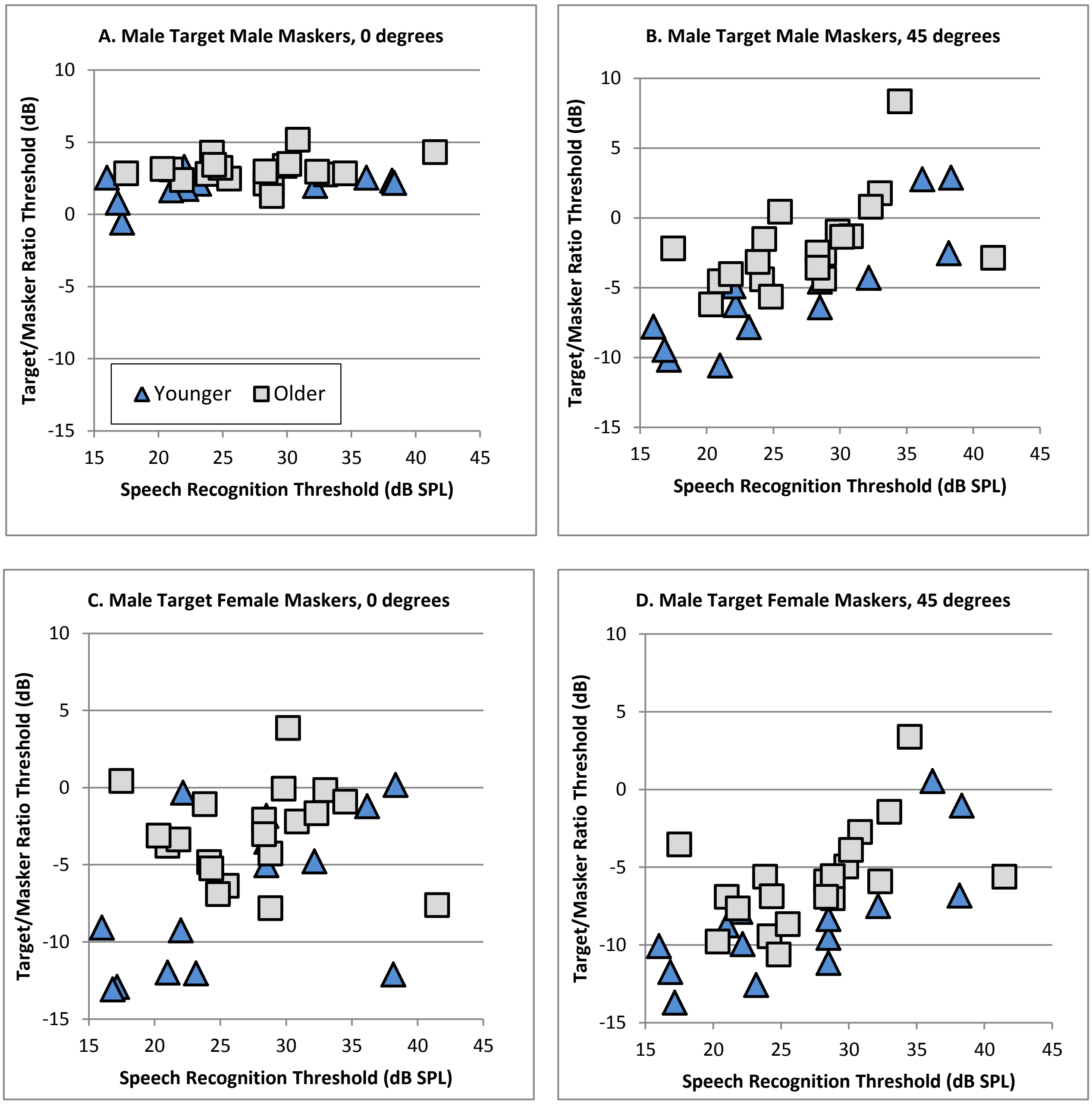

FIGURE 2. Relationships between Target/Masker Ratio and Speech Recognition Thresholds for older listeners (squares) and younger listeners (triangles) in four different combinations of target gender, masker gender, and spatial configuration.

TABLE 2. Correlation values between ITD threshold and target/masker ratio thresholds in eight representative spatial configurations. . Correlations marked with a single asterisk (*) are statistically significant at level of p<.05. All significance values are based on a sample size of n=17.

Target	*Male*	*Male*	*Male*	*Male*	*Female*	*Female*	*Female*	*Female*
Masker	*Male*	*Male*	*Female*	*Female*	*Female*	*Female*	*Male*	*Male*
Spatial Separation	*0 º*	*45 º*	*0 º*	*45 º*	*0 º*	*45 º*	*0 º*	*45 º*
Correlation	.201	.490*	.311	.291	.522*	.549*	.223	.329
Significance (n=17)	.439	.046	.224	.257	.031	.022	.390	.197

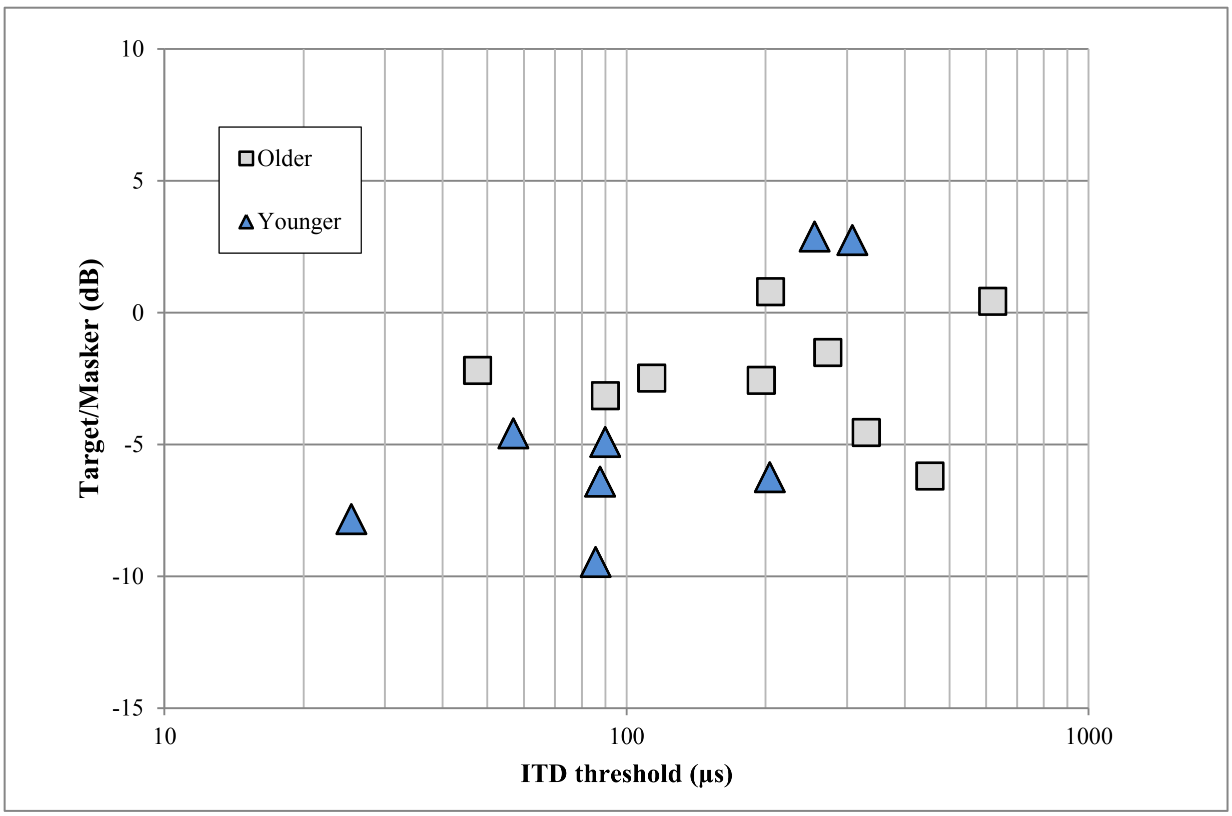

FIGURE 3. Relationships between ITD thresholds and Target/Masker Ratio for a male target with male maskers in the 45 degree separation condition (r = .49, p < .05). Square symbols indicate the scores of the older listeners and triangles indicate the scores of the younger listeners.

ACKNOWLEDGMENTS

We are grateful to the many participants, both Veterans and non-Veterans, who volunteered their time to this study. Assistance in the areas of data collection, database management, and compliance assurance was rendered by Erin Conner, Patrick Tsukuda, and Jaime Carlson. The work was supported by the Department of Veterans Affairs Rehabilitation Research and Development Service (Career Development Award C4963W) and the National Institutes of Health's National Institute for Deafness and Communication Disorders (R01 DC011828).

REFERENCES

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

Green, D. M. and Swets, J. A. **(1974)**. *Signal Detection Theory and Psychophysics*. (Wiley and Sons, New York).

Kidd G., Mason C. R., Richards V. M., Gallun F. J., and Durlach N. I. **(2008)**. "Informational masking," in *Auditory Perception of Sound Sources*, edited by Yost W. A., Popper A. N., and Fay R. R. (Springer Science+Business, New York).

Marrone N, Mason CR, Kidd G Jr. **(2008)** "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am., 124(5): 3064-3075.

Neher T., Laugesen S., Jensen N.S., Kragelund L. **(2011)** "Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities?" J. Acoust. Soc. Am., **130**, 1542-1558.

Noble W., Gatehouse, S. **(2006)** "Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial and Qualities of Hearing Scale (SSQ)," Int. J. Audiol. **45**, 172-181.