www.czasopisma.pan.pl www.journals.pan.pl

ARCHIVES OF ACOUSTICS
36, 1, 3–14 (2011)

DOI: 10.2478/v10168-011-0001-4

Sentence Recognition in the Presence of Competing Speech Messages Presented in Audiometric Booths with Reverberation Times of 0.4 and 0.6 Seconds

Kim S. ABOUCHACRA[(1)], Janet KOEHNKE[(2)], Joan BESING[(2)], Tomasz LETOWSKI[(3)]

[(1)] *American University of Beirut – Medical Center*
Department of Otolaryngology-Head and Neck Surgery
P.O. Box 11-0236, Beirut, Lebanon
e-mail: ks05@aub.edu.lb

[(2)] *Montclair State University*
Department of Communication Sciences & Disorders
Montclair, New Jersey, U.S.A.

[(3)] *U.S. Army Research Laboratory*
RDRL-HR/B-459
Aberdeen Proving Ground, Maryland, U.S.A.

(*received July 14, 2010; accepted January 21, 2011*)

This study examined whether differences in reverberation time (RT) between typical sound field test rooms used in audiology clinics have an effect on speech recognition in multi-talker environments. Separate groups of participants listened to target speech sentences presented simultaneously with 0-to-3 competing sentences through four spatially-separated loudspeakers in two sound field test rooms having RT = 0.6 sec (Site 1: $N = 16$) and RT = 0.4 sec (Site 2: $N = 12$). Speech recognition scores (SRSs) for the Synchronized Sentence Set (S^3) test and subjective estimates of perceived task difficulty were recorded. Obtained results indicate that the change in room RT from 0.4 to 0.6 sec did not significantly influence SRSs in quiet or in the presence of one competing sentence. However, this small change in RT affected SRSs when 2 and 3 competing sentences were present, resulting in mean SRSs that were about 8–10% better in the room with RT = 0.4 sec. Perceived task difficulty ratings increased as the complexity of the task increased, with average ratings similar across test sites for each level of sentence competition. These results suggest that site-specific normative data must be collected for sound field rooms if clinicians would like to use two or more directional speech maskers during routine sound field testing.

Keywords: sound field testing, reverberation, speech recognition.

1. Introduction

Speech perception testing is routinely conducted in clinical audiology settings via earphones (e.g., speech-threshold and word-discrimination testing) (Dirks *et al.,* 1972; Mendel, Danhauer, 1997). Unfortunately, earphone testing does not provide enough practical insights to predict how an individual will function in listening situations that more closely represent everyday communication environments. A more natural approach to assessing speech perception is through sound field testing, where the influences of noise and room acoustics on listener perception can be examined and estimates of difficulty with understanding conversation in background noise can be made (Taylor, 2003).

Sound field tests conducted at audiology clinics are traditionally limited to one or two loudspeakers located at a 1 meter distance from the listener. Often, a loudspeaker is positioned at 0°, 45°, or 90° azimuth in respect to the direction faced by the listener. The signal and masking noise, if any, are played from the same loudspeaker or the masking noise is played from a second loudspeaker typically located at the rear of the listener. In some clinics, masking noise is presented from several loudspeakers surrounding the listener to create a uniform background noise floor in the room. Sound field tests conducted under these conditions are intended to test the listener's ability to recognize speech in a quiet environment as well as in the presence of a single directional noise or uniform masking noise. The masking noise is usually speech spectrum noise, pink noise or multi-talker babble.

The masking noise conditions traditionally used for sound field testing in audiology clinics are neither the most prevalent nor most detrimental listening conditions a person may encounter in real world situations. In a typical listening situation, a person is affected by several competing sound sources (talkers) sending various messages. Even more importantly, the message the person needs to attend to changes from moment to moment and may originate from a different talker and direction over time (Shinn–Cunningham, Best, 2008).

The real world listening conditions described above seem to be very difficult and confusing. However, listeners typically display an amazing ability to cope with such conditions. For example, people have an acute ability to tune their attention to just one voice from a multitude of voices arriving from various directions during social gatherings. This phenomenon is usually referred to as the *cocktail party phenomenon* and has been extensively researched for years (e.g., Cherry, 1953; Bronkhorst, 2000). However, people also differ widely in this ability and traditional sound field tests do not specifically account for it.

In order to make both normative and clinical tests more predictive about human auditory behavior in real world situations, many research studies have been conducted using various types of signal (e.g., nonsense syllables, spondaic words, monosyllabic words, sentences) and maskers (e.g., flat-spectrum noise, speech-spectrum noise, modulated noise, babble, words, sentences) arriving from a variety of locations in space (e.g., Begault, Erbe, 1994; Dirks, Wilson, 1969; Festen, Plomp, 1990; MacKeith, Coles, 1971; Peissig, Kollmeier, 1997; Ericson *et al.*, 2004). This body of research has lead to a better understanding of the peripheral and central auditory processes involved in differentiating individual messages in a mixture of others, as well as to promising models for explaining the bottom-up and top-down components involved in the cocktail party phenomenon (*For reviews*, see Bronkhorst, 2000; Bregman, 1990; Schneider *et al*, 2007; Shinn–Cunningham, Best, 2008).

In recent years several new speech perception tests have been introduced that may allow clinicians to assess a patient's ability to localize sounds and hear speech signal in the presence of other localized distracters (e.g., other talkers present at distributed locations in the environment) (e.g., Killion *et al.*, 2004; Nilsson *et al.*, 1994). A special class of these tests is multi-channel sentence tests intended to test a patient's ability to hear and recognize specific speech strings in the presence of a number of competing speech messages arriving from other nearby locations. These tests assess a patient's ability to function in multi-talker environments such as social gatherings (e.g., cocktail parties, diners, and garden parties), public transportation venues, board and conference room meetings, stock exchange room activities, tactical operation centers (TOCs), and military and emergency action sites. Examples of test materials well-suited for use in multi-channel tests include the Coordinate Response Measure (CRM) test (Moore, 1981; Bolia *et al.*, 2000), the Synchronized Sentence Set (S^3) test (Abouchacra, 2000; Abouchacra *et al.*, 2001; Abouchacra *et al.*, 2009), and the Grid test (Cooke *et al.*, 2006). The new tests often utilize several loudspeakers (sound sources) distributed across a sound field test room to present target and masker speech messages, with individual loudspeakers having no predetermined role to be either the source of signals or the source of masking sounds. These tests have been used successfully to assess the effectiveness of multi-channel audio displays and speech enhancement technologies (active noise reduction and simulated spatialized sound presentation), and their potential use in clinical settings has been suggested (Abouchacra *et al.,* 2001; 2009; Cooke *et al.,* 2006; Ericson *et al.,* 2004).

The most natural implementation of multi-talker speech recognition tests in a clinical setting would be to adapt existing sound field test rooms for this kind of testing. However, such rooms differ from clinic to clinic both in size and in reverberation time (RT), which varies from about 0.3 to 0.6 sec for medium-sized audiometric booths. While it is well known that rooms with longer RTs decrease speech intelligibility, the effect of rooms with measured RTs less than 0.6 sec (e.g., audiometric test booths) has been only marginally studied. It is generally accepted that results of speech tests conducted in audiometric booths are equivalent if background noise levels are below certain limits, and the same test procedures, speech materials, and sound sources (loudspeakers) are used regardless of some possible small differences in RT However, such an assumption causes a concern, when considering multi-talker listening tasks. It is unclear whether existing sound fields found in audiology clinics are uniform enough to assure that the data obtained in different clinical facilities are directly comparable. The presence of several sound sources simultaneously emitting their signals results in mutual masking of one sound source production by the other, triggers different patterns of room vibrations and reflections, and may affect the intelligibility of what is heard by a listener in the room. As a result, differences in patients' performance across sites may occur because of known variations in the acoustic characteristics of the sound field audiometric booths (ASHA, 1991; Rochlin, 1993).

The purpose of the present study was to determine if differences in the RT between typical sound field test rooms used in audiology clinics have an effect on speech communication in multi-talker environments. To make this assessment, a listener's ability to correctly recognize target sentences in the presence of 0-to-3 masking sentences was determined in two different sound field test rooms having measured RTs of 0.4 sec and 0.6 sec. Sentences from the Synchronized Sentence Set (S^3) pool (Abouchacra, 2000) were presented through loudspeakers in the rooms to create the multi-talker environment, and speech recognition scores were measured and compared across test site, using young listeners with normal hearing.

2. Materials and methods

2.1. Participants

Twenty-eight listeners (aged 18–45 yr) from two test sites were paid for their participation in the study [Site 1: $N = 16$; Site 2: $N = 12$]. All were native speakers of English and had normal hearing thresholds in each ear (≤ 15 dB HL at octave frequencies from 0.25 to 8 kHz).

2.2. Stimuli

Target and competing sentence stimuli were randomly selected from the Synchronized Sentence Set (S^3) (Abouchacra, 2000). The S^3 includes 2304 ten-syllable sentences recorded by four male talkers (9216 total sentences); thus, the maximum number of simultaneous talkers that can be presented is 4. Each S^3 sentence consists of four token words embedded in a carrier phrase: "[NAME], *write the number* [NUMBER] *on the* [COLOR] [OBJECT]". Four NAME tokens (Troy, Nate, Mike, Ron), eight NUMBER tokens (1, 2, 3, 4, 5, 6, 8, 9), eight COLOR tokens (red, blue, green, pink, brown, black, white, gray), and nine OBJECT tokens (ball, cup, fork, key, kite, spoon, square, stair, star) can be randomly inserted into the carrier phrase. Onset times and intensity levels of the token words and carrier phrase were designed to be equivalent across sentences; thus, when sentences are presented simultaneously, the listener perceives them as synchronous in time and equivalent in loudness. Further information about the S^3 recordings and companion software can be found in Abouchacra *et al.* (2009).

2.3. Rooms and equipment

Testing was conducted in two medium-sized rooms used for audiological sound field testing at two different sites. The test spaces selected for the study were judged as representative of the ends of the range of reverberation times found in typical sound field test rooms. Specifically, the room at Site 1 was a $2.8 \times 3.4 \times 2.5$ m space having a reverberation time (RT_{60}) of 0.6 sec, and the room at Site 2 was a $2.8 \times 2.8 \times 2.1$ m space having a RT_{60} of 0.4 sec. Both test rooms had almost the same critical distance (CD) of about 0.35 m and were very similarly furnished. Since also the same type of loudspeaker was used in both spaces, the only meaningful differences between the two rooms were RTs and room dimensions.

A personal computer, custom software, and Tucker Davis Technologies (TDT System II) four-channel hardware were used to control simultaneous presentation of one target sentence and 0-to-3 masker sentences. The sound sources were four loudspeakers (Realistic, Minimus 7) located 1 m away from the listener at fixed positions of $\pm 135°$ and $\pm 45°$ azimuth (ear level), where $0°$ azimuth represents the location directly in front of the listener and negative azimuths indicate positions to the listener's left. This experimental setup and loudspeaker arrangement resulted in the sentences from Talkers 1, 2, 3, and 4 being presented to participants from loudspeakers arranged in a square configuration at fixed locations of -45, $+45$, -135, and $+135°$ azimuth, respectively. Such an arrangement does not favor any single direction and makes all angular separations identical. This was an important consideration since the focus of the study was on speech recognition and not on the influence of spatial configuration.

Loudspeakers in each test room were calibrated and matched in terms of overall level when measurements were made at the listener's head position (listener absent). Pink noise was used as a calibration signal. With the microphone placed at the location corresponding to the center of the listener's head, the output signals generated by all four loudspeakers were equalized from 0.2 to 9 kHz (1/3 octave bands) to be equal within ± 1 dB. Finally, forty S^3 stimuli spoken by each male talker were directed through their respective loudspeaker. Measured speech peaks for the S^3 tokens in the sentences averaged 65 dB SPL (± 1 dB) for each loudspeaker, when measured at the eardrum of an acoustic manikin (KEMAR) positioned in the participant's test location. No adjustments were made to the level of the overall signal (target plus masker(s)) during the experiment (i.e., the target-to-masker ratio worsened as the number of talkers increased).

2.4. Procedure

Although a different group of participants served as listeners at each test site, the stimuli, equipment and procedures were identical. The listener's task was to record token words from target sentences that were presented in isolation (0 competing sentences) or in conjunction with 1, 2 or 3 competing sentences. Participants were instructed to listen for the target sentence in each trial, as indicated by the token NAME "*Troy*". They were told that the target sentence could originate from any of the loudspeakers (talkers) and would be presented alone or concurrently with up to 3 competing sentences.

Target sentences were presented 160 times to the participant, with 40 trials in each level of message competition (0, 1, 2, and 3). A single trial lasted 10 seconds (3 sec to present the sentence(s) and 7 sec for the participant to manually record his/her response). A recording of a female voice (presented simultaneously through all four loudspeakers) was used to announce the trial number and cue the participant for listening. The participant was asked to sit upright with his/her back against the chair and look straight ahead (no head restraint was used). After each target sentence was presented, the participant registered the NUMBER, COLOR, and OBJECT tokens of the target sentence on a closed-set response form. For example, a correct response for the target sentence, "***Troy***, *write the number **3** on the **black ball***", is shown in Fig. 1. Participants heard each target sentence once. No feedback was given.

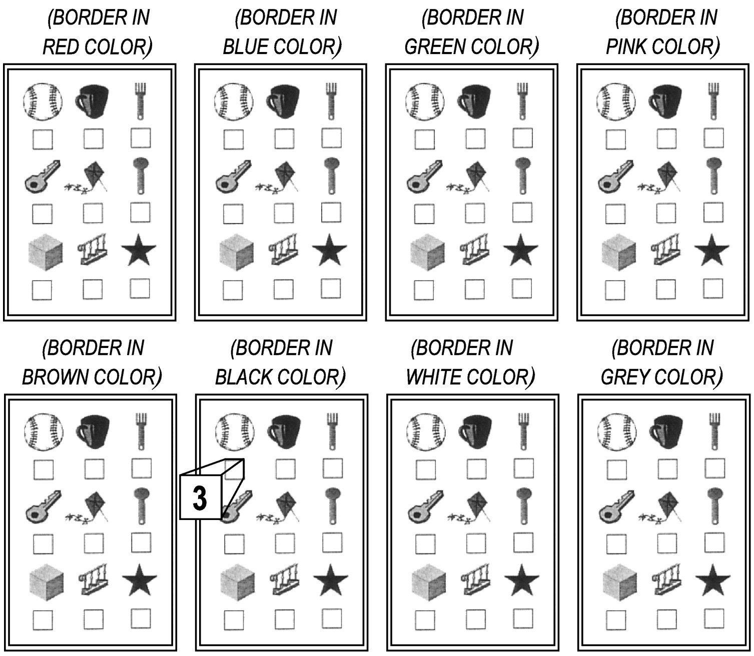

Fig. 1. Closed-set response form for recording S^3 test responses. The form is in black-and-white, with the exception of the borders around each block of objects; the color is indicated in this figure above each block. The participant writes the target NUMBER in the response box below the target OBJECT within the appropriately colored border; borders around each block of objects represented the COLOR token. This figure shows a correct response for the T-message, "***Troy***, *write the number **3** on the **black ball***".

At the end of each competing sentence condition (0, 1, 2, and 3), each participant rated perceived task difficulty on a 5-point rating scale (1 = easy to understand, 2 = slightly difficult to understand, 3 = moderately difficult to understand, 4 = very difficult to understand, 5 = impossible to understand).

2.5. Experimental design

The dependent variable in this study was *sentence recognition score* (SRS), defined as the total percentage of target sentences recorded correctly over a block of test trials. Two independent variables were manipulated in the experiment, including the type of test room (between-subject variable) and the number of concurrent competing sentences (within-subject variable). All listeners completed one block of 40 test trials in each level of sentence competition (i.e., 0, 1, 2, or 3 competing sentences), for a total of 160 trials/participant. The order of concurrent competing sentences always increased from 0 to 3 sentences. During a test block, 10 target sentences from each loudspeaker/talker were presented in random order. The NUMBER, COLOR, and OBJECT tokens were randomly chosen, with replacement, for each trial. However, the software program prevented the same tokens from being presented by more than one source (talker) during a given trial. The testing of all four levels of sentence competition in a given room (i.e., 160 trials) required approximately one hour, including rest periods and a block of training trials. The purpose of the training was to familiarize participants with the sentence structure, number of competing sentences to expect, and procedure for recording responses.

3. Results

Speech recognition ability was scored separately for each participant as the percentage of sentences that were correctly identified in each listening mode and referred to as the *speech recognition score* (SRS). To get full credit for the sentence, the participant's response needed to be free of any errors. Mean SRSs and corresponding standard deviations (SD) for participants tested at each site and listening condition are summarized in Fig. 2.

A two way analysis of variance (ANOVA) with one fixed factor (SITE; Site 1, Site 2) and one repeated measure factor (C-MESSAGE; 4 levels) indicated statistically significant differences for the main effects of test site [$F(1, 26) = 6.27$; $p = 0.019$], sentence competition [$F(3, 78) = 273.88$; $p < 0.0001$], and their interaction [$F(3, 78) = 3.17$; $p = 0.029$]. More specifically, the ANOVA results indicated significantly better performance by participants at Site 2 (RT = 0.4 sec) than Site 1 (RT = 0.6 sec) in the most difficult sentence competition conditions (i.e., 2 and 3 masker sentences) and no significant differences between sites for test conditions with 0 and 1 competing sentences. All levels of the sentence competition factor were significantly different at $p < 0.04$ level or less.

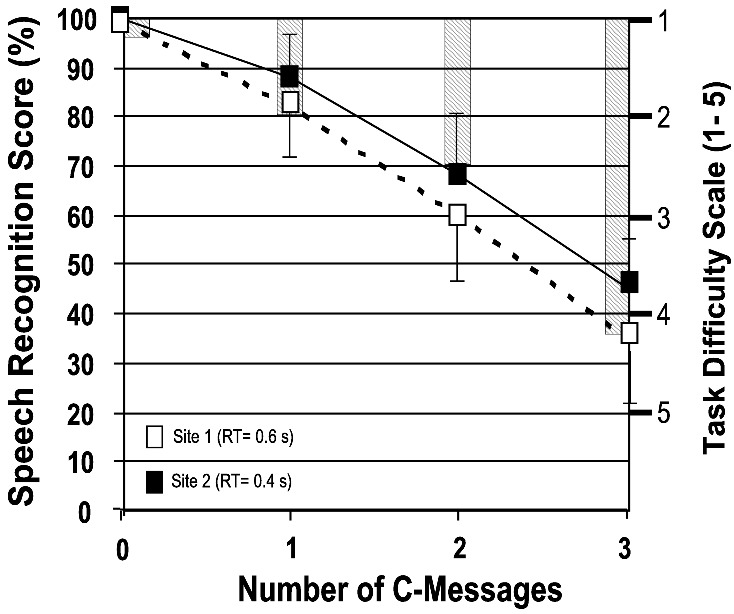

Fig. 2. Mean speech recognition scores and perceived listening difficulty ratings presented as a function of the number of competing messages. Speech recognition scores are represented by the line graphs (Site 1: open squares with dotted line; Site 2: solid square with solid line). The dashed bars represent mean listening difficulty judgments for all participants from both sites.

An analysis of participants' ratings of listening difficulty in each *listening mode x C-message* condition revealed no significant difference in ratings between test sites ($p > 0.05$). Statistically significant differences in subjective ratings were found, however, for C-message condition [$F(3, 51) = 555.71$; $p < 0.0001$]. These data indicated that perceived task difficulty increased as the complexity of the listening task increased from easy (0 C-messages) to difficult (3 C-messages).

4. Discussion

Reverberation can disrupt a listener's ability to selectively attend to target signals in a natural environment. For example, reverberation is known to corrupt interaural time and level difference cues, and in the case of speech stimuli, reverberation smears the harmonic structure of speech (especially in high frequencies), decreases the depth of amplitude modulation in speech, blurs speech onsets/offsets and formant movements, and fills spectrotemporal gaps that are normally present in speech (e.g., Cooke, 2006). The resulting consequences of reverberation on multi-talker listening are increased overlap-masking of sound sources (talkers) (Nabelek *et al.*, 1989), and reduced effectiveness of features that can be used by listeners to differentiate a target speech signal from competing speech messages (Darwin, Hukin, 2000a; 2000b).

Data obtained in the present study revealed that in two rooms with reverberation times of $\mathrm{RT}_{60} = 0.4$ sec and $\mathrm{RT}_{60} = 0.6$ sec and having the same critical distance of 2.3 m, the S^3 SRS values were not affected by the RT when sentences were presented in isolation (0 competing messages). This good performance likely occurred because sound sources (loudspeakers) were relatively close to the listener (1 m); thus, the direct sound dominated over any early and late reflections. This, however, was not the case when multiple concurrent sentences were present in the rooms.

Statistically significant differences in speech recognition scores were observed between the test sites when 2 or 3 competing messages were present in the listening environment, with better performance occurring in the room with a measured RT_{60} of 0.4 sec. As shown in Fig. 2, the group mean SRSs observed in the $\mathrm{RT}_{60} = 0.4$ sec room were by 8% and 10% better than those observed in the $\mathrm{RT}_{60} = 0.6$ sec room for the 2 and 3 competing sentence conditions, respectively. Although it was not found to be statistically significant, a 6% difference in mean SRSs was also noted between test sites for the 1 competing sentence condition. Because the experimental design was identical across test sites, the increasing differences in speech recognition scores between groups tested at different tests sites (as more competing sentences were added) likely resulted from the differences in room reverberation times. In other words, increased reverberation exacerbated any overlap-masking effects caused by increased sentence competition and significantly decreased target sentence recognition. Adding another masking speech source in the room can be visualized as adding a stream of strong directional (although uncorrelated) reflections.

Reported data indicate that the S^3 test used in this study was sensitive enough to capture the effects of $\mathrm{RT}_{60} = 0.4$ sec and $\mathrm{RT}_{60} = 0.6$ sec on speech recognition scores obtained by young listeners with normal hearing listening to speech signals with multiple speech maskers. The task difficulty ratings suggest a gradual subjective increase in task difficulty as the number of competing messages increased.

It is important for the reader to recognize that similar performance results may not likely occur if broadband noise were presented through loudspeakers as the competing signal. Previous research has demonstrated that different types of masking will be dominant as a function of test signal, which could lead to differences in performance. When broadband noise is used as the competing signal, *energetic masking* dominates (Kidd *et al.*, 2008). Energetic masking refers to masker energy that overwhelms the energy of the target signal at the level of the basilar membrane or auditory nerve (i.e., by activating similar regions of the auditory periphery) (Durlach *et al.,* 2003). On the other hand, when speech is used as the competing signal, *informational masking* dominates (Kidd *et al.,* 2005; Shinn–Cunningham, 2005). Informational masking is a broad term used to describe interference beyond energetic masking that is believed to involve higher-level processes such as those used in attention, memory, cognition, perceptual grouping, and source segregation, among others (Durlach *et al.,* 2003;

Kidd *et al.,* 2008; Schneider *et al.,* 2007). Informational masking is believed to be vital to the understanding of the cocktail party problem (Brungart, 2001). In the case of the multi-talker task used in this study, both types of masking are influencing performance to some extent; however, it is likely that informational masking is the dominating type when the S^3 pool is used, as the recording method used in the construction of these sentences was designed to create a high degree of informational masking when multiple sentences are presented simultaneously.

5. Conclusions

The findings of this study indicate that sound field test room differences in RT from 0.4 to 0.6 sec do not significantly influence speech recognition scores of listeners with normal hearing in quiet or in the presence of one speech distracter. However, differences in sound field reverberation may affect speech recognition scores when two or more independent speech distracters are present. These results suggest that site-specific normative data must be available for the multi-talker stimuli, if clinicians would like to use two or more directional speech maskers for routine sound field testing and be able to exchange normalized data. Another possibility is to determine an uncertainty range for such type of data when use for clinical purposes.

Even more importantly, collected data indicate that differences in acoustic properties of the sound field test rooms that are currently used in the audiology clinics may affect some types of collected data. Thus, presented findings support a previously expressed need (e.g., Rochlin, 1993) for more stringent requirements and more detailed operational conditions for sound field rooms, which may be beneficial for both repeatability and reliability of traditional single-loudspeaker and newly developed distributed-loudspeakers speech recognition tests.

Acknowledgments

The authors would like to thank Mr. Tuyen Tran for designing the initial version of the custom software used in conjunction with the S^3 corpus. This work was supported by the U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

References

1. Abouchacra K.S. (2000), *Synchronized sentence set* (S^3) [*Compact Disk* #3], U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, Maryland.
2. Abouchacra K.S., Breitenbach J., Mermagen T, Letowski T. (2001), *Binaural Helmet: Improving speech recognition in noise with spatialized sound*, Human Factors, **43**, 584–594.

3. Abouchacra K.S., Letowski T., Besing J., Koehnke J. (2009), *Clinical application of the Synchronized Sentence Set* (S^3), Proceedings of the 16th International Congress of Sound and Vibration, Kraków.

4. American Speech-Language-Hearing Association (1991), *Sound Field Measurement Tutorial. Working Group on Sound Field Calibration of the Committee on Audiological Evaluation,* American Speech-Language-Hearing Association **Supplement Jan**, 25–38.

5. Begault D.R., Erbe T. (1994), *Multichannel spatial Auditory display for speech communications*, Journal of the Audio Engineering Society, **42**, 819–826.

6. Bolia R., Nelson W., Ericson M., Simpson B. (2000), *A speech corpus for multitalker communications research,* Journal of the Acoustical Society of America, **107**, 1065–1066.

7. Bregman A.S. (1990), *Auditory Scene Analysis: The perceptual organization of sound*, MIT Press, Boston.

8. Bronkhorst A.W. (2000), *The Cocktail Party Phenomenon: A review of research on speech intelligibility in multiple-talker conditions*, Acustica – Acta Acustica, **86**, 117–128.

9. Brungart D.S. (2001), *Informational and energetic masking effects in the perception of two simultaneous talkers*, Journal of the Acoustical Society of America, **109**, 1101–1109.

10. Cherry E.C. (1953), *Some experiments on the recognition of speech, with one and with two ears*, Journal of the Acoustical Society of America, **25**, 975–979.

11. Cooke M. (2006), *A glimpsing model of speech perception in noise*, Journal of the Acoustical Society of America, **119**, 1562–1573.

12. Cooke M., Barker J., Cunningham S., Shao X. (2006), *An audio-visual corpus for speech perception and automatic speech recognition*, Journal of the Acoustical Society of America, **120**, 2421–2424.

13. Darwin C., Hukin R. (2000a), *Effectiveness of spatial cues, prosody, and talker characteristics in selective attention*, Journal of the Acoustical Society of America, **107**, 970–977.

14. Darwin C., Hukin R. (2000b), *Effects of reverberation on spatial, prosodic and vocal-tract size cues to selective attention*, Journal of the Acoustical Society of America, **108**, 335–342.

15. Dirks D.D., Wilson R.H. (1969), *The effect of spatially separated sound sources on speech intelligibility*, Journal of Speech and Hearing Research, **12**, 5–38.

16. Dirks D.D., Stream R.W., Wilson R.H. (1972), *Speech audiometry: earphone and sound field*, Journal of Speech and Hearing Disorders, **37**, 62–76.

17. Durlach N.I., Mason C.R., Kidd Jr. G., Arbogast T., Colburn H.S., Shinn–Cunningham B.G. (2003), *Note on informational masking*, Journal of the Acoustical Society of America, **113**, 1984–1987.

18. Ericson M.A., Brungart D.S., Simpson B.D. (2004), *Factors that influence intelligibility in multitalker speech displays*, The International Journal of Aviation Psychology, **14,** 313–334.

19. Festen J.M., Plomp R. (1990), *Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing*, Journal of the Acoustical Society of America, **88**, 1725–1736.

20. Kidd Jr. G., Arbogast T.L., Mason C.R., Gallun F.J. (2005), *The advantage of knowing where to listen*, Journal of the Acoustical Society of America, **118**, 3804–3815.

21. Kidd Jr. G., Mason C.R., Richards V.M., Gallun F.J., Durlach N.I. (2008), *Informational masking*, [in:] *Auditory Perception of Sound Sources*, Yost W.A., Popper A.N., Fay R.R. [Eds.], pp. 143–190, Springer Science + Business Media, New York.

22. Killion M.C., Niquette P.A., Gudmundsen G.I., Revit L.J., Banerjee S. (2004), *Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners*, Journal of the Acoustical Society of America, **116**, 2395–2405.

23. MacKeith N.W., Coles R.R.A. (1971), *Binaural advantages in hearing of speech*, Journal of Laryngology and Otology, **85**, 214–232.

24. Mendel L.L., Danhauer J.L. [Eds.], (1997), *Audiologic evaluation and management and speech perception assessment*, Mendel, Singular Publishing Group, Inc., San Diego.

25. Moore T. (1981), *Voice communication jamming research.* AGARD Conference Proceedings 331: Aural Communication in Aviation, 2:1-2:6, Neuilly-Sur-Seine, France.

26. Nabelek A., Letowski T.R., Tucker F.M. (1989), *Reverberant overlap- and self-masking in consonant identification*, Journal of the Acoustical Society of America, **86**, 1259–1265.

27. Nilsson M., Soli S.D., Sullivan J. (1994), *Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and noise*, Journal of the Acoustical Society of America, **95,** 1085–1099.

28. Peissig J., Kollmeier B. (1997), *Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners*, Journal of the Acoustical Society of America, **101**, 1660–1670.

29. Rochlin G. (1993), *Status of sound field audiometry among audiologists in the United States*, Journal of the American Academy of Audiology, **4**, 2, 59–68.

30. Schneider B.A., Li L., Daneman M. (2007), *How competing speech interferes with speech comprehension in everyday listening situations*, Journal of the American Academy of Audiology, **18**, 478–591.

31. Shinn–Cunningham B.G. (2005), *Influences of spatial cues on grouping and understanding sound*, Proceedings of Forum Acusticum, 29 August – 2 September 2005.

32. Shinn–Cunningham B., Best V. (2008), *Selective attention in normal and impaired hearing*, Trends in Amplification, **12**, 283–299.

33. Taylor B. (2003), *Speech-in-noise tests: How and why to include them in your basic test battery*, The Hearing Journal, **56**, 40–46.