Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 1021–1032

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Speech intelligibility for different spatial configurations of target speech and competing noise source in a horizontal and median plane

Edward Ozimek *, Jędrzej Kociński, Dariusz Kutzner, Aleksander Sęk, Andrzej Wicher

Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University in Poznan, Umultowska 85, 61-614 Poznan, Poland

Received 5 February 2013; received in revised form 5 June 2013; accepted 13 June 2013
Available online 24 June 2013

Abstract

The speech intelligibility for different configurations of a target signal (speech) and masker (babble noise) in a horizontal and a median plane was investigated. The sources were placed at the front, in the back or in the right hand side (at different angular configurations) of a dummy head. The speech signals were presented to listeners via headphones at different signal-to-noise ratios (SNR). Three different types of listening mode (binaural and monaural for the right or left ear) were tested. It was found that the binaural mode gave the lowest, i.e. 'the best', speech reception threshold (SRT) values compared to the other modes, except for the cases when both the target and masker were at the same position. With regard to the monaural modes, SRTs were generally worse than those for the binaural mode. The new data gathered for the median plane revealed that a change in elevation of the speech source had a small, but statistically significant, influence on speech intelligibility. It was found that when speech elevation was increased, speech intelligibility decreased.

Keywords: Speech intelligibility; Speech-in-noise test; Spatial perception; Monaural and binaural perception

1. Introduction

In natural acoustic environments speech often coexists with signals generated by other sound sources. Therefore, communication is often made difficult, because speech is masked by other sounds (speech, traffic noise, music etc.). However, the auditory system is capable of separating out signals coming from different directions and extracting information of interest. Many experiments on speech intelligibility measurements for different configurations of sources have been carried out so far. For example, Bronkhurst and Plomp (1988) investigated the effect of interaural time delay (ITD) and acoustic headshadow on binaural speech intelligibility in noise. Recordings were made of speech reproduced in front of a manikin; and of noise emanating from seven angles in the horizontal plane ranging from 0° (frontal) to 180° in steps of 30°. Freyman et al. (2001) determined the extent to which the perceived separation of speech and interference improves speech recognition in the free field. The target talker was always presented from a loudspeaker directly in front (0°). The interference was either presented from the front or from both a right loudspeaker (60°) and a front loudspeaker, with the right leading the front by 4 ms. In the experiment carried out by Hawley et al. (2004), speech reception thresholds (SRTs) were measured for Harvard IEEE sentences presented from the front in the presence of one, two, or three interference sources. Moreover, four types of interferer were used: other sentences spoken by the same speaker, time-reversed sentences of the same speaker, speech-spectrum shaped noise, and speech-spectrum shaped noise, modulated by the temporal envelope of the sentences. In the research of Kociński and Sęk (2005), the speech intelligibility in the presence of one or two statistically independent speech-shaped noise sources varying in configuration was investigated. Litovsky (2005) tested children between the ages of 4 and 7, and adults, in free field

* Corresponding author. Address: Institute of Acoustics, Adam Mickiewicz University, Ul. Umultowska 85, 61-614 Poznan, Poland. Tel.: +48 618295133; fax: +48 618295123.
E-mail address: ozimaku@amu.edu.pl (E. Ozimek).

0167-6393/$ - see front matter
http://dx.doi.org/10.1016/j.specom.2013.06.009

for speech intelligibility. The target speech was presented from the front (0°); speech or modulated speech-shaped-noise competitors were either in front or on the right (90°). Brungart and Iyer (2012) showed that listeners with normal hearing are able to efficiently extract information from better-ear glimpses that fluctuate rapidly across frequency and across the two ears.

A general aim of these experiments was to investigate a spatial release from masking (also called spatial unmasking or spatial suppression). It turned out that the speech intelligibility depended among others on the mutual spatial configuration of the speech and masker sources. Moreover, it was found that for a given signal-to-noise ratio (SNR) the speech intelligibility was higher when the target and masker sources were spatially separated than in the case when they were collocated. The spatial unmasking is related to the influence of the listener's head on the propagation of the signal that creates an acoustical shadow (called also the head shadow effect) leading to an interaural level difference. When the speech and noise sources are spatially separated, the listener is able to take advantage of this difference that leads to changes in SNRs in the respective ears. Moreover, the interaural time (phase) differences play an important role in the spatial unmasking, which can be partially interpreted in terms of the binaural masking level difference (Kociński and Sęk, 2005). This effect (that originally was related to the detection of tones presented in noise) was generalized to speech perception and was called the binaural intelligibility level difference (Peissig and Kollmeier, 1997), which incorporates both binaural and monaural components of auditory processing (Garadat and Litovsky, 2006; Hawley et al., 2004; Lin and Feng, 2003).

Many investigators have analysed configurations in which the target speech source was placed directly in front of the listener and the azimuth of the disturbing source was varied, e.g. 0° (frontal) to 180° in steps of 30° (Bronkhurst and Plomp, 1988), 0° and 60° (Freyman et al., 1999) or 0°, 45°, 90° (Drullman and Bronkhorst, 2000). A scenario in which speech and masker sources are spatially distributed was also used in a recent study by Allen et al. (2008) that concerned auditory streaming and spatial speech unmasking. In the quoted study considerable release from masking was demonstrated when two noise sources were located symmetrically at azimuths of -30 deg and $+30$ deg, respectively.

The purpose of the present study was to determine the SRT (i.e. SNR yielding 50% speech intelligibility) for different configurations of the speech and masker sources. The experiments were carried out in monaural and binaural listening modes, in the horizontal plane (outlined by the surface that cuts through the head at ear level) (experiment 1) and in the median plane (outlined by surface that splits the head into the left and the right halves, i.e. sagittal plane in anatomical coordinate system) (experiment 2). In Fig. 1, the above mentioned planes, the spatial configurations of speech and noise considered in the experiments as well as angles describing direction of an incoming sound (i.e. azimuth and elevation), are schematically presented.

Since a vast majority of previously carried out experiments focused mainly on the influence of the masker azimuth, while the speech azimuth was kept at 0°, within a framework of this study speech azimuth was considered (the experiment 1) for the following listening modes: the monaural-left ear, the monaural-right ear and binaural (for details see Section 2). The experiment 2 analysis the novel aspect of the speech intelligibility referring to the effects in median plane. In this case the speech azimuth was kept constant, while the speech source elevation and the masker azimuth were modified and the signals were presented monaurally and binaurally. There are several situations in which the sources of speech signals are situated higher than listeners' heads. For example, in most public places, like stations, airports or churches, speech sources are located above our heads. Therefore it seems to be fairly important to check speech intelligibility experimentally when speech sources are located above the head.

The Polish Sentence Test (PST) (Ozimek et al., 2009) was used as the target speech material for the first time in such a study. It is worth adding that the rationale for doing this study is both scientific and practical. For example, comparison of the intelligibility data obtained for the monaural-left and right ear modes in acoustically adverse conditions can provide relevant information on the location of speech and noise sources to get optimal hearing conditions, especially for patients with a unilateral hearing loss. Furthermore, in different human environments, speech and noise sources are often distributed each others in a complex way. The results of the present study could be helpful in defining the optimal spatial distribution of those sources from the speech intelligibility improvement point of view.

2. Materials and methods

2.1. Stimuli: speech and masking noise

The PST, which reflects the basic features of the Polish language, consists of 25 different lists, each containing 20 sentences. The test is characterized by a large number of fricatives and limited number of vowels. Due to this, a relatively high level of energy in the frequency range above 5 kHz can be noticed, and a high variability in amplitude envelope occurs in comparison with other languages. When presented in a background of the so-called babble noise masker, the lists of sentences produce relatively steep intelligibility functions, i.e. functions that link the probability of correct response to SNR. A large slope of intelligibility function at the SRT point implies low inter- and intra-variability across the lists and, consequently, a high precision in SRT determination can be obtained. Using such a test it is possible to detect subtle differences in speech intelligibility for different measurement conditions, which might be difficult to obtain for test materials producing less steep