Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 66 (2015) 82–90

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

The interaction between speech intelligibility task and non-auditory tasks

Jędrzej Kociński *

Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University in Poznań, Poland

Received 27 March 2014; received in revised form 29 September 2014; accepted 8 October 2014
Available online 12 October 2014

Abstract

Many research on various aspects of speech perception and intelligibility has been conducted so far. However, the ability of human mind to perform multiple tasks simultaneously is still a challenging topic. Perceptual load theory states that people first decide what activity is most important after which available perceptual resources are allocated. Less important activities are conducted effectively only if remaining resources are sufficient. It should be emphasized that researchers studying divided attention focused mainly on activities requiring vision. This paper focuses on interaction between non-auditory tasks (distracters) and a speech intelligibility task. A very sensitive method based on speech reception threshold (SRT) measurement was used instead of standard speech intelligibility determination. This sensitivity is a result of a very steep psychometric function (ipso facto small standard deviation) which is not possible to achieve for standard methods. Therefore, this property can be used to investigate even slight impact of external conditions on speech perception. Two experiments examining the interactions between these two tasks were conducted. Despite the use of such sensitive method, in none of the experiments speech intelligibility decreased due to distracter. However, the percentage of correct responses to the distracter when subjects were performing speech intelligibility tests decreased compared to the situation when they focused only on the distracter.

Keywords: Speech intelligibility; Distracter; Interaction; Load theory

1. Introduction

Speech is one of the most important, but also most complex signals found in nature. Nevertheless, despite many varying factors that describe the speech signal, people are able to communicate without problems. Equally surprising is how much redundant information is in the speech signal. Thus, a high speech intelligibility is possible even in conditions where the energy of interfering noise is several times greater than the energy of the speech signal (Duquesnoy, 1983; Festen and Plomp, 1986; Gustafsson and Arlinger, 1994; Kocinski and Sek, 2005; Ozimek et al., 2013; Plomp, 1976). It must be emphasized that in a natural environment, beside a speech signal, many other signals that involve other senses and attention occur simultaneously. These so-called distracters can influence our ability to process a speech signal correctly. A compelling example, which became the direct inspiration for research in this area, is the effect of driving a car on the perception of speech signal (e.g. during phone conversations), and how the simultaneous perception of speech can affect driving (Brown et al., 1969; Dressel and Atchley, 2008; Levy et al., 2006; Treffner and Barrett, 2004). The activities, which hinder the speech perception are generally called distracters in these papers.

It seems reasonable to state that the speech signal is the most important signal for human beings. This may be the reason that a specialized mode of our mind or a brain state

* Address: Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University in Poznan, ul. Umultowska 85, 61-614 Poznan, Poland.
E-mail address: jedrzej.kocinski@amu.edu.pl

http://dx.doi.org/10.1016/j.specom.2014.10.002
0167-6393/

(mental processing) focused on the perception of speech was found (Liberman et al., 1967; Moore, 2003). This mode is a mechanism responsible for the interpretation of sounds which are detected to be "speechlike" by the human mind (Parker et al., 1986). It facilitates communication and is used unconsciously, as soon as a perceived signal meets the criteria to be defined as a speech signal. Evidence for the existence of speech mode was presented by Remez et al. (1981). They investigated the so-called sinusoidal speech, namely a signal consisting of three tones, whose frequencies and levels were corresponding to the first three formants of speech. The frequencies were changing in the same way as the frequency of the speech formants. In this way, some sentences were generated and presented to the listeners. The subjects were asked about the material they just heard. Most of them responded that they had heard a strange kind of sounds associated with the electronic music or science-fiction movies. The same experiment was carried out on people who were informed that they would hear distorted sentences. This time subjects were able not only to say that they heard speech-like sounds, but also were able to understand the sentences. The explanation is that the second group of subjects were able (according to the given suggestions) to use a mode that allowed them to analyze the perceived signal as speech (speech mode). Similar results were obtained by Gardzielewska (2007) who also investigated synthetic speech. In her experiments the sinusoidal speech was also used. It turned out that even when the speech synthesis uses only 3 sinusoidal components, the intelligibility of words in sentences was about 70%, although the resulting signal was very little like real speech. Obtaining such high intelligibility can be explained just by the existence of speech mode which was used by the listeners. Moreover, once the speech mode is activated, it is very difficult to "switch it off" (Stevens and House, 1972). In most cases, it is activated automatically and unconsciously by the perception of signals with specific properties (Moore, 2003). Therefore, it seems reasonable to state that the existence of the special speech mode can be reflected in the speech intelligibility task presented simultaneously with distracters from other modalities.

It must be emphasized that not only noise or reverberation can affect the speech intelligibility, but also signals from other senses. One of the most convincing phenomenon that supports the assumptions on the interactions between modalities was shown by McGurk and McDonald (1976). It was shown that the human mind processes information reaching him through different senses at the same time. The impact of information reaching the mind through the sense of sight (or hearing) on the information coming through the sense of hearing (or vision) is called the audiovisual merge. In this experiment the authors prepared the audiovisual samples so that the visual track presented the person pronouncing the syllable while the audio was simultaneously presenting another syllable. With such a treatment the listeners perceived the syllable other than heard and saw. Most listeners were not informed of the conflict of audio and video track. Although these results are difficult to interpret because of the complexity of the brain and the processing made, they confirm, that the influence of senses on each other exists, and audiovisual merge may have, in certain situations, a significant effect on speech intelligibility. In later studies, the existence of audiovisual merge was confirmed by Nicholls et al. (2004). Moreover, when inconsistent stimuli are presented to a subject they can be perceived as completely disjunctive ones with no interactions between them, however a more complicated interaction can also be found, namely one can become more dominant and can attract an attention, e.g. Colavita effect (Colavita, 1974), in which an audio stimulus (tone) and visual stimulus (flash) were presented to a listener. The task was to push a different button for visual and audio stimulus as fast as the subject noticed the stimulus. Most of the samples were only in one modality (audio or visual), but some of them were bimodal. When only an audio stimulus was presented, a subject had no problem to accomplish the task, and generally for this stimulus his/her reaction was faster than for visual one, however when two stimuli were presented simultaneously, all subjects pushed the button for visual stimulus. This effect was confirmed by Koppen and Spence (2007) and Sinnett et al. (2007, 2008).

Furthermore, attention experiments show more interactions between these modalities. Duncan et al. (1997) investigated the ability of subjects for noticing different items (spoken or written) in a series of the same items. They showed that the percentage of correct answers decreases when stimuli from both modalities are presented together. Moreover, when two streams from one modality (vision or audio) were presented to a subject the lowermost score was obtained when both stimuli were presented simultaneously and was increased when there was a time shift between them, while for interaction between modalities (one stimulus from vision and one from audio) the time shift did not bring about any change in the score. Sometimes, however, the information from one modality can also help in perception of information from the other sense. Best et al. (2008) showed that a visual stimulus that suggested the direction of the audio stimuli to be presented or time when the audio stimuli would be presented increases the percentage of correct answers. Moreover, when both visual prompts (time and direction) were presented, the score was the highest.

It must be emphasized that human mind is able to process only a certain amount of information simultaneously. The maximum amount of information processed at the same time is called the capacity of attention (Plack, 2010). Thus, one can assume that the performance of more than one activity, or perception of more than one signal simultaneously, even from different senses, can lead to the possibility of exceeding the maximum capacity of attention and some parts of analyzed information may be lost, or incorrectly interpreted. The ability of the human mind to analyze multiple streams of information is an object of experiments of the perceptual load theory (Francis, 2010;