The Connected Ear
Besser, J.

2015

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 23. Apr. 2024
INTRODUCTION

People talk. And they talk a lot. An average of 16,000 spoken words per person per day were measured in a study of Mehl, Vazire, Ramírez-Esparza, Slater en Pennebaker (2007). While observations regarding the magnitude of daily verbal output differ between studies, the communicative nature of humans is not in doubt. Obviously, one purpose of spoken communication is the transfer of essential factual information. However, with 16,000 words per day, there is room for much more than that. Spoken communication is a core component of human interaction and an important factor in building interpersonal relationships with far-reaching meaning for an individual’s quality of life. For example, it has been found that a person’s happiness is related to the amount of their serious in-depth conversations (Mehl, Vazire, Holleran & Clark, 2010). As a consequence, losing the tools to engage in conversations can have serious psychosocial implications, whether the production (e.g., Chaves et al., 2012; Cruice, Worrall, Hickson & Murison, 2003; Jacobson et al., 1997) or the perception (e.g., Bess, Lichtenstein, Logan, Burger & Nelson, 1989; Dalton et al., 2003; Heine & Browning, 2002; Kramer, Kapteyn, Kuik & Deeg, 2002; Strawbridge, Wallhagen, Shema & Kaplan, 2000) of speech is regarded. A framework introduced by Mattys, Davis, Bradlow en Scott (2012) describes the negative influence of irregularities in speech production and speech perception on spoken communication as a result of their degrading effects on speech-signal quality, causing adverse conditions for communication. This thesis focusses on the perceptual side of spoken communication. Specifically, while spoken communication in face-to-face conversations is a multimodal type of interaction, this thesis considers the listener’s auditory-only perception and processing of speech. Previous research has linked speech-understanding abilities to a variety of listener-internal factors, including the listener’s age, hearing status, cognitive functioning, and auditory temporal processing. Age as such is unlikely to be a determinant for speech understanding, but is assumed to represent relevant processing mechanisms that change with increasing age (Houtgast & Festen, 2008). One aim of this thesis was to identify cognitive (Chapter 2 through Chapter 5) and auditory temporal (Chapter 4 and Chapter 5) processing abilities that can (partially) explain the age effect observed
in speech understanding. Furthermore, while the influence of certain cognitive and auditory temporal-processing abilities on speech understanding has often been examined in isolation, this thesis aims to answer the question how the individual influences combine in speech understanding, taking into account the listener’s age and hearing status.

Next to listener-internal factors, the previously mentioned framework by Mattys et al. (2012) also describes environmental influences that can adversely affect spoken communication because of signal degradation. In alignment with this notion, attention is also given to the fact that people typically socialize in a variety of listening situations that can vary considerably with regard to their acoustic properties. One of the factors determining the character of a listening situation is the presence of sounds interfering with the target speech, also called maskers, which create an adverse listening condition. The negative effect of masking sounds on speech understanding is an accepted fact. However, less is known about the influence of the type of masker on the employment of the aforementioned auditory and cognitive processing mechanisms in speech understanding. This thesis aims to further the understanding of the interactions between masker and listener-internal processing abilities in speech understanding.

1.1 Hearing loss – its prevalence and implications

Hearing loss can negatively affect spoken communication with direct or indirect consequences for a person’s quality of life and social well-being. Given that hearing loss is one of the worldwide most prevalent chronic and disabling conditions (Stevens et al., 2013; World Health Organization, 2008), this constitutes a serious concern of public health. People with a hearing loss create significantly higher health care costs than people without a hearing loss, because of an increased number of medical consultations and treatments directly linked to the hearing impairment (Nachtegaal et al., 2010). Furthermore, self-reported sick leave due to mental distress is increased for people with a hearing loss (Kramer, Kapteyn & Houtgast, 2006) and increased levels of longer sick leave have also been found for people with decreased hearing ability in noise, mediated by a higher need for recovery after work (Nachtegaal, Festen & Kramer, 2012). Hearing loss at low frequencies (250 and 500 Hz) and self-perceived hearing difficulties have been observed to lead to early retirement in middle aged men and women (Helvik, Krokstad & Tambs, 2013).

In the vast majority of cases, the development of a hearing loss is closely aligned with aging, a phenomenon also called age-related hearing loss or presbycusis (Gates & Mills, 2005). Prevalence estimates (World Health Organization, 2012) indicate that a third of all adults 65 years or older suffer from disabling hearing loss. This esti-
mate is in agreement with data presented by Stevens et al. (2013) who found that the prevalence of hearing loss was less than 10% for people at the age of 50 years, 20% for people at the age of 60 years, more than 40% for people at the age of 70 years, and more than 50% for people at the age of 80 years. These numbers illustrate that not only the prevalence of hearing loss, but also the rate of hearing decline increases with increasing age, which was also observed by Gates en Cooper (1991). Consequently, the number of people living with a hearing loss increases with the number of elderly in the world’s population. And the number of elderly increases because of increasing longevity. It is thus of great relevance to find ways to secure speech-communication skills over the adult lifespan into high ages, which in turn requires knowledge of the mechanisms driving successful speech understanding. Therefore, this thesis examines the effects of aging on speech understanding for people without and people with a sensorineural hearing loss.

1.2 Threshold and Suprathreshold Influences on Speech Understanding

Age has been found to be a determining factor in the ability to understand speech (Committee on Hearing Bioacoustics and Biomechanics, 1988). According to CHABA (1988), on an individual basis, the age effect can be related to age-dependent alterations in one or several of the following three processing domains: peripheral, central, and cognitive. Peripheral influences denote any effect related to deficits in the auditory periphery, i.e., the outer ear, the middle ear, and the inner ear. An example of peripheral issues are audiometric threshold elevations, accounting for the attenuation component in auditory processing as described by Plomp (1978). Influences of audiometric thresholds on speech understanding are also referred to as threshold factors in the current thesis. They constitute the most apparent source of difficulties in speech understanding in older people (e.g., Divenyi, Stark & Haupt, 2005; Schmiedt, 2010; van Rooij & Plomp, 1992).

However, age-related difficulties in speech understanding can occur independent of elevated audiometric thresholds (e.g., Dubno, Dirks & Morgan, 1984; Pichora-Fuller & Souza, 2003; Tun, McCoy & Wingfield, 2009) in all of the previously mentioned processing domains. For example, spectral processing, such as frequency selectivity (e.g., Hopkins & Moore, 2011; Strelcyk & Dau, 2009) has typically been linked to processing abilities of the inner ear (peripheral). Temporal resolu-

---

1 Following Humes en Dubno (2010), the term "speech understanding" will be used throughout this thesis to denote the ability to acoustically perceive and semantically interpret a speech signal to the degree that the listener can reproduce the signal’s verbal content. Occasionally, the terms "speech perception", "speech reception", or "speech comprehension" may also be used in the same meaning, if no other meaning is explicitly stated.
tion (e.g., George et al., 2007; Humes, Kewley-Port, Fogerty & Kinney, 2010) and binaural processing of interaural time and level differences (e.g., Glyde, Buchholz, Dillon, Cameron & Hickson, 2013; Goverts & Houtgast, 2010) on the other hand have been related to decreased auditory central processing, i.e., processing deficits in the central auditory pathways, which consist of a system of auditory nerve fibers, several nuclei and colliculi, and the auditory cortex (Ehret & Romand, 1997). Independent of their physiological origin, perceptual effects of auditory functions beyond audiometric thresholds are referred to as suprathreshold factors in this thesis.

The thesis investigates how speech understanding is influenced by monaural auditory suprathreshold processing in the temporal domain, both for binaural understanding of target speech which is spatially separated from the masker (Chapter 4) and for monaural understanding of speech which is not spatially separated from the masker (Chapter 5). Speech contains large amounts of temporal information (Rosen, 1992) and consequently, auditory temporal processing has been found to be influential for speech understanding (Humes, 2013; Humes, Kidd & Lentz, 2013; Pichora-Fuller & Souza, 2003). There are different kinds of temporal information in a speech signal which all contribute to speech understanding independent of binaural temporal processing, i.e., temporal envelope (e.g., Fogerty, 2011; Purcell, John, Schneider & Picton, 2004; Ruggles, Bharadwaj & Shinn-Cunningham, 2012), temporal fine-structure (e.g., Ardoint, Sheft, Fleuriot, Garnier & Lorenzi, 2010; Jackson & Moore, 2013; Ruggles et al., 2012), and periodicity (Summers & Leek, 1998; Vongpaisal & Pichora-Fuller, 2007) cues. Each type of temporal information is associated with different kinds of information in the speech signal (see Chapter 4) and the ability to process these different types of information can be reduced independently. Furthermore, temporal processing also plays a role for the ability to make use of temporal fluctuations in background noise (George, Festen & Houtgast, 2006).

Cognitive influences denote effects of individual abilities in higher-order processing that are not directly linked to the auditory periphery or auditory central pathways. Similar to auditory suprathreshold factors, the perceptual effects of cognitive processing can occur without auditory threshold elevations. They are therefore also referred to as nonauditory suprathreshold factors in this thesis. Cognitive contributions to speech understanding have been observed early on both in complex listening situations with multitalker displays that require selective attention to understand a specific talker (Broadbent, 1958, cited in Arlinger, Lunner, Lyxell & Pichora-Fuller, 2009; Cherry, 1953) and in the minimalistic context of word recognition with masked phonemes, which requires top-down processing skills to fill in the missing information (Warren, 1970). However, only in the last two decades has the interest in cognitive contributions to speech under-
Cognitive hearing science emerged as a steady and prominent subdomain of hearing research (Arlinger et al., 2009).

Cognitive abilities influence speech understanding at all ages. For example, the ability to fill in missing linguistic information affects speech understanding in noise in adult listeners independent of age (see Chapter 2). Nonetheless, cognitive influences on speech understanding are more pronounced or more easily detected in older listeners, which may partly be explained by the connection of sensory and cognitive declines in aging as evidenced in the Berlin Aging Study (Baltes & Lindenberger, 1997). However, importantly, also people without clinical hearing loss (i.e., sensory decline) can experience difficulties understanding speech in complex listening situations, related to cognitive processing abilities (e.g., Houtgast & Festen, 2008; Kiessling et al., 2003; Pichora-Fuller, Schneider & Daneman, 1995; Schneider, Pichora-Fuller & Daneman, 2010; Wingfield & Tun, 2007).

The types of cognitive abilities that have been found to influence speech understanding are manifold. This thesis focuses on verbal working memory (WM) capacity (Chapter 2, Chapter 3, Chapter 5), linguistic closure abilities (Chapter 2 through Chapter 5), general information processing speed (Chapter 2), and general cognitive ability (Chapter 4). Here, linguistic closure denotes the ability to decode fragmentary linguistic information and correctly identify its content. For this purpose, the missing parts need to be filled in by top-down processing. Thus, rather than referring to a form of crystallized knowledge, such as vocabulary, linguistic closure refers to the ability to apply crystallized linguistic knowledge. The listed cognitive factors' influences on speech understanding in specific adverse listening conditions are investigated, both for younger listeners with good hearing, for older listeners with a hearing loss and for older listeners without a hearing loss.

### 1.3 Adverse Listening Conditions

In addition to the source and the receiver of a speech signal, another essential part of verbal communication is what is often referred to as the ‘channel’ or the ‘transmission’ of the signal. Even with a “perfect” source and a “perfect” receiver, communication can be very difficult if the channel is noisy, which is also called an adverse listening condition. Examples of adverse listening conditions are having a conversation next to a running jackhammer, at a pop concert, or at a children’s party. Sometimes, the term ‘adverse condition’ is used to denote any factor (internal or external to the listener) that complicates listening (Mattys et al., 2012). In contrast, in this thesis the term is used to represent factors influential for spoken communication that are external to both the target speaker and the listener. For older people
especially, it has been found that they have difficulties understanding speech in such conditions, related to the threshold and suprathreshold processing abilities mentioned above. In the research presented in this thesis, adverse conditions were implemented by the use of disturbing background sounds during assessments of speech understanding. The amount of overlap of the masker with the target speech was varied along the dimensions of temporal properties, spatial location, semantic information, and voice features.

1.4 THE HISTORICAL AND SCIENTIFIC CONTEXT OF THIS THESIS

The research presented in this thesis is far from being the first work studying the mechanisms underlying speech understanding. The thesis builds on a tradition of research on auditory perception and speech understanding that reaches back more than a century. An important prerequisite for later hearing research was the acknowledgment that not only physical objects themselves, but also the perception thereof can be studied in an exact and systematic manner. An early attempt to formulate the laws determining the relationship between the physical and the psychological realm was made by Fechner (1860) in his book “Elemente der Psychophysik” (Elements of Psychophysics). Indeed, psychophysical experiments have played an important role in research on hearing and speech understanding and continue to do so. The essence of such experiments is that they systematically vary one parameter of interest and test how this influences performance, thereby examining specific functions and limits of auditory perception. The audiogram is a basic, but meaningful example of this approach, which reflects both the general intent of psychoacoustics to exclude any effects of factors other than the controlled manipulation and a preference for working with physically well-defined signals such as sinusoidal tones. Traditionally, the focus of psychoacoustic research has been to understand the functioning of lower-level sensory processes in the peripheral auditory system and the physiological structures that are involved in the process (Neuhoff, 2004, Chapter 1).

Another important development in the second half of the 19th century was the invention of the telephone by A.G. Bell, which triggered extensive research on hearing and speech transmission and perception at AT&T Bell Laboratories and other sites, including the development of standardized tests and materials (e.g., Fletcher & Steinberg, 1929; see also Wilson & McArdle, 2005 for a review). Demands of telephonic communication guided research in the years after World War II (Studdert-Kennedy & Whalen, 1989) and eventually, investigations on the effect of noisy transmissions on speech intelligibility led to the

formulation of several models that aim to predict speech intelligibility based on the acoustic environment, i.e., the Articulation Index (AI; ANSI, 1969; French & Steinberg, 1947), the Speech Intelligibility Index (SII; ANSI, 1997), and the Speech Transmission Index (STI; Steeneken & Houtgast, 1980).

These standardized models of speech intelligibility do not take into account factors internal to the listener other than audiometric thresholds. However, in parallel to the formulation of intelligibility indexes, other areas of hearing research evolved and it was detected early on that cognitive factors, such as attentional mechanisms, play a role in speech understanding in natural listening situations. An example of this is the legendary work by Cherry (1953) describing the "cocktail party problem", which refers to the challenge of listening to one talker in a multi-talker setting. Also a commentary by Myklebust (1949) and a schematic overview of the involved processes presented in H. Davis (1964; reprinted in Pichora-Fuller & Singh, 2006) attest an early awareness of the entanglement of psychology and audiology and the role of memory, attention, and language in speech understanding. Contributions of cognitive factors to speech understanding have continued to engage hearing researchers and are still a topic of investigation, as this thesis exemplifies.

The development of the parallel threads of research on speech perception, some of which were briefly described in the preceding paragraphs, has gone in a cyclical fashion, much like described in Kuhn’s (1962) reflections on paradigm shifts and the principles of scientific development. Depending on the zeitgeist and technological possibilities, ideas flourished or were neglected under certain periods of time, but they keep coming back when "the time is ripe". Across these developments, Reinier Plomp (2001, p. 2) in his book "The Intelligent Ear" identified four types of scientific biases that have affected the history of hearing research:

"(a) the dominance of sinusoidal tones as stimuli, (b) the predilection for a "microscopic" (as opposed to a macroscopic) approach, (c) emphasis on the psychophysical (rather than the cognitive) aspects of hearing, and (d) focus on stimuli abstracted from the "dirty" conditions of everyday listening."

According to Plomp (2001), bias (a) has its roots in the 19th century’s development of physics, where Fourier’s detection of the composition of periodical functions led hearing researchers to focus on the perception of sinusoidal tones and to view hearing as a passive process determined by the functioning of a series of bandpass filters. Bias (b) builds on the assumption that the functioning of the auditory system is the sum of the functioning of its subsystems and thus that it is sufficient to study the perception of isolated elements, such as pure-tones or phonemes, to explain the mechanisms of speech perception.
Plomp (2001) describes bias (c) as a preference in hearing research to study auditory aspects of hearing and listening, based on the notion that auditory and cognitive processing are two separate systems that can be studied in isolation. Bias (d) illustrates that researchers have a preference for testing in listening conditions that differ from the auditory scenes people experience in real life, which is related to bias (b) and an attempt to keep the experimental setups as “clean” and controlled as possible.

These biases reflect a general historical contrast in science between reductionism – Gestaltism on the one hand and gestaltism on the other hand. According to the principles proposed by reductionism, the process of speech understanding could be fully explained by knowing the involved mechanisms in isolation, such as the perception of acoustic features – e.g., frequency, duration, harmonic structure and signal level – and the perception of the smallest elements of speech (phonemes). Gestaltism on the other hand proposes a more holistic functioning of the perceptual system, where auditory objects are perceived as wholes that cannot be described as a mere sum of their subparts. According to the biases formulated by Plomp – especially bias b – reductionist approaches have dominated science, including hearing research. This circumstance may be related to scientific beliefs, but practical reasons may have been equally relevant. A few decades ago, conducting hearing research was equipment intensive and even simple psychoacoustic experiments required large hardware setups. Technological advances made old types of data collection easier and new types possible, which in turn boosted the further development of existing and new ideas (see also Neuhoff, 2004, Chapter 1). For example, the introduction of neuro-imaging techniques and other physiological measures has opened up new ways to study interactions of peripheral and cognitive influences in speech understanding (e.g., Anderson, White-Schwoch, Parbery-Clark & Kraus, 2013; Peelle, 2012). It should be noted that physiological and behavioral studies of cognitive mechanisms as such do not grant a holistic scientific approach, but the increased awareness of higher-order influences on speech understanding has led to a steady increase in the number of publications that studied auditory and cognitive influences on speech understanding in combination (Pichora-Fuller & Singh, 2006). Thus, while the different threads of hearing research continue to develop separately, an increasing degree of entanglement of the threads can also be observed. The research presented in this thesis affiliates with this recent course of integrated, macroscopic studies and tries to overcome the biases formulated by Plomp (2001) to some extent by 1) explicitly focusing on cognitive aspects of speech understanding, 2) acknowledging that only combinations of auditory and cognitive factors can explain speech-understanding abilities, and 3) including listening conditions that attempt to resemble realistic listening situations, e.g., by using...
interfering speech as a masker and by spatially separating the target and the interfering sound sources.

1.5 OUTLINE OF THE THESIS

Each of the chapters from Chapter 2 through Chapter 5 presents one speech-understanding experiment. The experiments differ in their focus regarding characteristics of the participants (age, hearing status), masker signal (steady-state noise, temporally modulated noise, interfering speech), masker location (same as or different from target), and/or the assessed suprathreshold factors (cognitive, linguistic, auditory temporal). The main purpose of the studies presented in Chapter 2 and Chapter 3 was to design and evaluate appropriate tests for linguistic closure abilities and working memory span, respectively, that were later applied in the experiments of Chapter 4 and/or Chapter 5.

Chapter 2 presents an experiment that examined contributions of age, education, sex, linguistic closure abilities, WM span, and processing speed to speech understanding in steady-state and temporally modulated noise. Participants were from a wide age range (18–78 years) and had no reductions in pure-tone hearing acuity >= 20 dB HL for octave frequencies at 500–4000 Hz. The experiment’s primary purpose was to further develop the existing text reception threshold (TRT) test that has been designed to measure the modality-independent\(^3\) cognitive and linguistic abilities involved in speech understanding as measured by the speech reception threshold (Plomp & Mimpen, 1979). Based on suggestions from earlier research (Kramer, Zekveld & Houtgast, 2009), new test versions were developed with the aim to increase the shared variance of TRT and SRT by making the TRT more dependent on age, working-memory capacity, and processing speed. The experiment identified one of the revised TRT versions as the most suitable TRT test for the test battery employed in the remaining experiments of this project.

Chapter 3 formally introduces the Dutch WM span test developed within the thesis project and presents an experiment examining modality differences (reading vs. listening) in verbal WM capacity and associations of WM span in the different modalities to speech understanding in temporally modulated noise. Younger adults (19–35 years) without reductions in pure-tone hearing acuity >= 20 dB HL for octave frequencies at 125–8000 Hz participated in this study. The experiment described in Chapter 2

\(^3\) Note that linguistic influences on speech understanding have also been found within the auditory modality in studies evidencing phonemic restoration effects in noise-interrupted speech (e.g., Bashford, Riener & Warren, 1992; Baskent, Eiler & Edwards, 2010; Benard, Mensink & Baskent, 2014).
had uncovered some fundamental differences between the abilities measured by the TRT test and verbal WM span. Therefore, the main purpose of the study presented in Chapter 3 was to follow up these previous observations by performing a literature review that investigated the relationships of TRT and WM span, respectively, to SRTs in different masking conditions.

Chapter 4. Like the studies presented in the previous chapters, this study included listeners without clinically significant hearing loss, namely younger (18–27 years) and older (66–82 years) adults with normal audiometric thresholds (≤ 25 dB) in most of the speech range (250–3000 Hz). Unlike the earlier studies, this experiment made use of speech maskers that interfered semantically with the target speech and examined effects of target and masker locations. In addition, the test battery was extended with auditory temporal measures. The study’s focus was twofold: First, age group differences were assessed in SRTs for speech-in-speech listening with and without spatial separation and with or without voice differences between target and masker speech. Second, the experiment examined how speech understanding in these listening situations was influenced by age, audiometric thresholds, auditory processing of temporal envelope and fine-structure cues, general cognitive ability, and linguistic abilities.

Chapter 5 presents a study that examined most of the masking conditions examined in the previous chapters in combination with most of the previously examined predictors for speech understanding. In other words, speech understanding in steady-state and fluctuating noise-maskers and with an interfering talker was examined in dependence of age, audiometric thresholds, auditory temporal processing, WM capacity, general cognitive ability, and linguistic abilities. Effects of spatial separation cues were not examined in this study. Instead, the study served to investigate differences in influences of suprathreshold processing between older listeners without (51–70 years) and older listeners with (45–83 years) a sensorineural hearing loss. In addition, the study aimed to shed light on the question how distinct types of suprathreshold processing interact with specific adverse speech-understanding situations.

Chapter 6 provides a general discussion of the findings from the experiments conducted in this project.