Cognitive Hearing Science: The Role of a Working Memory System for Speech Understanding in Old Age

Jerker Rönnberg1,2, Mary Rudner1,2 and Adriana A. Zekveld1,2,3

Abstract

The chapter is about cognitive hearing science (CHS). CHS focuses on the interaction between bottom-up and top-down processing in speech understanding. This chapter will focus on the listening and comprehension stages of speech understanding for hearing-impaired persons under challenging listening conditions. One relevant conceptualization in this context is the Ease of Language Understanding model (ELU). The model and its key features will be briefly described, and some of the clinical implications will be addressed, especially those that pertain to old age. We address issues of (1) signal processing in hearing aids and working memory (WM) capacity, (2) long-term episodic memory consequences of hearing impairment, (3) ELU and ageing, and (4) WM training in old age.

Introduction

As articulated in recent publications, Cognitive Hearing Science is an emerging field that stresses the interplay between bottom-up and top-down processing in language understanding under adverse conditions (Akeroyd, 2008; Arlinger et al., 2009; Campbell et al., 2009; Edwards, 2007; Gatehouse et al., 2003, 2006; Rönnberg & Arlinger, 2008). Cognitive hearing science is a new, interdisciplinary field that focuses on how hearing-impaired and deaf people deploy cognitive resources to communicate in realistic, everyday situations. Crucial to the field are the conceptualizations of how the dynamic interplay between human cognition and auditory signal processing become manifest in the nervous system. The design of hearing instruments must take this interplay into account. Central to the area are how cognitive and linguistic resources change as a function of age, how these changes affect speech understanding, and how the interplay with a hearing instrument continues to affect the perceptual and also the cognitive functions involved.

A Working Memory System for Ease of Language Understanding (ELU)

Clinicians may consider applying several assumptions and predictions pertaining to the working memory (WM) system for Ease of Language Understanding (ELU) developed by Rönnberg et al. (2003; 2008; 2009a; 2009b; 2010). The predictions concern (a) interactions between signal processing in hearing aids and cognitive resources, (b) effects of hearing impairment on long-term memory, (c) effects of ageing on ease of language understanding, and (d) WM training and old age. Before we deal with these implications for the elderly hearing impaired person, we present a description of the model. The model is about the various effects of different listening conditions on perceptual and cognitive load during speech comprehension. These listening conditions may be defined in relation to a number of parame-

---

1 Linnaeus Centre HEAD, Swedish Institute for Disability Research, Linköping University, Sweden.
2 Department of Behavioural Sciences and Learning, Linköping University, Sweden.
3 Audiology/ENT & EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam.

Address correspondence to: Jerker Rönnberg, Department of Behavioral Sciences and Learning, Linköping university, S-581 83 Linköping, Sweden. Email: jerker.ronnberg@liu.se, Fax: +46 13 282145.
ters concerning, for example, external noise, social context, hearing impairment or signal processing in a hearing instrument. In particular, the model addresses adverse listening conditions where mismatch may occur between the input signal and the memory representations. Mismatch may be related to any of the above-mentioned parameters but in the present chapter we focus on phonological mismatch. The model takes into account that language comes in different modalities, for example, the spoken word, the written word and the signs of sign language. Written and signed language are not dealt with in this chapter (but see Rönnberg et al., 2010; Rudner et al., 2009a). However, even spoken language comes in different modalities: visual (as in lipreading), auditory, by means of tactiling, or multimodal, i.e. different combinations of these modalities (Rönnberg, 2003; Rönnberg et al., 2008). The multimodal nature of the model is important from an ecological perspective, as natural communication typically involves both auditory and visual information and furthermore it provides a global theoretical framework for understanding language comprehension and cognition in relation to the complex interplay between sensory modalities and language modalities.

Critically, the model assumes that speech-related information delivered in different sensory modalities is rapidly bound together in a so-called episodic buffer (Baddeley, 2000; Repovs & Baddeley, 2006; Rudner & Rönnberg, 2008) that we denote RAMBPHO, i.e., Rapid Automatic Multimodal Binding of PHONology (Rönnberg et al., 2010). This binding is assumed to occur around 150 ms after stimulus onset. The assumption is based on findings relating to the McGurk effect (McGurk & MacDonald, 1976) showing that auditory and visual information in the speech signal are bound together in the nervous system at this point in time (Campbell, 2008). The McGurk effect is the phenomenon that arises when two different syllables are presented in the auditory and visual channels giving rise to the percept of a third syllable that has not been presented (McGurk & MacDonald, 1976; Campbell, 2008). For example, an auditory “ba” accompanied by a visual “ga” is heard as “da”.

Somewhere between 150–200 ms the RAMBPHO-extracted information is assumed to make contact with a corresponding phonological representation in long-term memory, by which the lexicon is accessed. This assumption is based on findings relating to mismatch negativity, which is the automatic electrophysiological brain response to discriminable change in auditory stimulation (Näätänen, 2008).

The precision of the phonological representation is an important parameter related to interindividual differences in phonological processing (Andersson, 2002). The time window for lexical access may last up to 300–400 ms (Poeppe, 2008), and its duration is an important feature of the model for predicting interindividual differences. If the phonological information extracted from the signal is insufficiently precise, the probability of a mismatch between the incoming information and the phonological representations in long-term memory increases. According to the ELU model, when mismatch arises, the implicit (unconscious) cognitive processing involved in language understanding under advantageous conditions is no longer sufficient for achieving comprehension, thus processing becomes explicit (conscious).

Explicit processing concerns the utilization of cognitive resources in order to repair misunderstandings and infer missing information in the chain of spoken tokens in a dialogue. It is about retrospective disambiguation and prospective prediction about what is to come in the dialogue. Explicit processing capacity may by critically dependent on the executive resources that coordinate storage and processing operations in WM according to a long tradition of WM models (Baddeley, 1986; Baddeley & Hitch, 1974; Rönnberg et al., 2008). In early models of WM (e.g. Baddeley, 1986; Baddeley & Hitch, 1974), the function of the episodic buffer, i.e. binding of information in different sensory modalities and from different sources to create a unified cognitive representation, was considered to be part of the central executive, a component of WM that includes executive functions and is considered responsible for overseeing WM processing. However, accumulating research shows that executive processing is not a necessary feature of the episodic buffer (Baddeley et al., 2009 Rudner & Rönnberg, 2008). Thus, whereas RAMBPHO processing is implicit and rapid, explicit processing is typically deliberate and slow, using a time window of several seconds (Stenfelt & Rönnberg, 2009).

It is assumed that mismatches occur more often for persons with hearing impairment than without, even when speech perception is aided by hearing instruments. This is because hearing instruments, although they may include a number of different technologies aimed at combating the specific difficulties experienced by persons with hearing impairment, cannot fully restore hearing function (e.g., Edwards, 2007). Consequently, on average, even when hearing loss is compensated by a hearing aid, a person with hearing impairment...
is more likely than a person without to be dependent on explicit cognitive resources during speech understanding in noise. This may in its turn be related to increased effort as reflected by physiological measures (Kramer et al., 1997), and by a hearing loss related increase in the need for recovery after work, and also by a higher incidence of distress related sick leave from work (Kramer et al., 2006; Kramer, 2008; Nachtegaal et al., 2009).

**Signal Processing in Hearing Aids and WM Capacity**

We have manipulated the probability of phonological mismatch during speech processing for speech in several studies. In a study by Foo et al. (2007), the pre-experimental compression release settings were different from the settings used in an experimental aid. In other words, participants with hearing loss were tested with unfamiliar hearing aid settings that differed from the ones they were used to. This caused a phonological mismatch effect across the board (see also Cox & Xu, in press). The mismatch effect was examined by assessing the relation between aided speech recognition in noise and the performance on a complex WM test. The WM test applied was the reading span test (for a review see Daneman & Merikle, 1996; first used in a hearing impairment context by Rönnberg et al., 1989). The rationale behind this approach is that if WM capacity is related to speech understanding performance in certain conditions, apparently, these conditions do impose a demand on these WM processes. Foo et al. observed that WM capacity was correlated with speech understanding performance in both stereotypical (Hagerman, 1997; 2002) and more naturalistic sentence conditions (Swedish version of HINT sentences; Hälgren et al., 2006).

Later studies have shown that – after 9 weeks of experience with a certain kind of compression release setting, and then speech recognition in noise testing with either the same or a different setting – it is especially in the different, mismatching conditions, that there is dependency on WM (Rudner et al., 2009b). After 18 weeks of experience of new compression settings including 9 weeks experience of slow-acting, linear compression and 9 weeks experience of fast-acting, non-linear compression – when the mismatch effects of the manipulation of the settings are presumably eliminated – there still remains an interesting effect of cognitive capacity. In challenging conditions (fast-acting compression and modulated noise), high capacity individuals (i.e., based on reading span) performed better on aided speech recognition in noise. Low capacity individuals benefited from the noise modulation only with slow compression settings (Rudner et al., 2009c; cf. Lunner & Sundewall-Thorén, 2007; Rönnberg et al., 2009a).

To the extent that noise modulation enhances speech recognition in noise performance, it also seems to gear the listener towards processing sentences such as the Hagerman sentences in a phonological (c.f., Mattys et al., 2009) rather than a semantic mode. This may in turn trigger a dependency on WM and explicit linguistic abilities since the modulations per se seem to signal a speech-processing mode that is qualitatively different than that pertaining in steady state noise, presumably because the individual is attempting to piece together the different speech segments in the dips of noise. This in itself demands a reliance on WM of the sort tapped by the reading span task, and involves storing old segments while processing and perceiving new ones, and then trying to synthesize them into what was actually uttered. Furthermore, if processing is interrupted by successive mismatches, then it is even more crucial to have a capacious WM to counteract the implicit processing challenges encountered (Rönnberg et al., 2009a). This situation is compounded by the effects of age. As described above, interindividual differences in WM capacity may predict actual performance levels under certain conditions and WM capacity correlates with age (Foo et al., 2007).

In order to more closely examine the effects of age on the interaction between cognition and aided speech recognition in noise performance, we performed a reanalysis of an existing data set (Foo et al., 2007). We observed that individuals with low WM capacity needed a 3 dB higher SNR in modulated noise than high-capacity individuals, but this applied only in the very old (mean age 76 year) and not in moderately old (mean age 64 year) listeners. The pattern held true across performance criteria (50%, 80%) (Rudner & Rönnberg, 2010). This indicates that for speech understanding in challenging conditions, cognitive capacity becomes increasingly important with advancing age, and that speech perception does not necessarily need to decline if cognitive capacity can be maintained (cf. Rönnberg, 2003). These results are important as aging is generally related to poorer speech perception in modulated noise conditions (Festen & Plomp, 1990; Larsby et al., 2008).
Other data from the memory literature suggest that elderly people unlike younger people do not derive the same relative benefit from semantic encoding instructions compared to shallow encoding instructions. They also show less activity in semantic processing regions of the brain such as the middle temporal lobe (Erber et al., 1989; Daselaar et al., 2003). Elderly people may therefore be more reliant on a phonological mode of processing and thus become more dependent on WM, which means that if they have a capacious WM, it still can be used to compensate for inability to reconstruct what was uttered in the dialogue (Rönnberg, 2003; Wingfield & Grossman, 2006). However, this issue is not settled as other data suggest that perception of sentences in noise in young and older listeners is facilitated similarly by semantic context (e.g. MacDonald, Davis, Pichora-Fuller & Johnsrude, 2008). The differences between the two kinds of data may lie in the comparison condition: a no-context condition versus a shallow encoding condition. In the latter case, we have a better control over what kind of processing the semantic processing is compared to. At any rate, the conclusion is that elderly persons can compensate if they have a capacious WM, and this may under some conditions be facilitated by semantic context.

Long-Term Episodic Memory
Consequences of Hearing Impairment

Episodic memory refers to the retrieval of time and space coordinates of encoding and retrieval of an experienced event, i.e., an episodic memory is the memory representation for a specific individual of a dated event that took place at a specific location. Semantic memory contains information that is abstract, impersonal, and not dated in time and space (Tulving, 1983). Phonological representations are linguistic; they belong to semantic memory, and phonetic contrasts are laid down early in life (i.e. before 6 months) according to the perceptual magnet theory (Kuhl et al., 1992).

Using a particular signal processing scheme during extended periods of intensive training may alter the way in which sounds are represented in semantic memory but perhaps in ways more subtle than the phonetic contrasts themselves (as indicated above). This may in its turn affect episodic memory functioning, as the route to encoding, storage, and retrieval of verbal information is via the precision and match of phonological representations.

A recent study (Rönnberg et al., 2009b), suggests that episodic memory performance in the elderly decreases as a function of increasing hearing impairment. The data were from a large database on cognitive aging (the Betula study, Nilsson et al., 1997), and the negative relationship with hearing impairment occurred despite the use of hearing aids. Structural equation modeling (SEM) was employed and indicated a significant relationship between two latent factors; one based on hearing impairment (across four frequencies), and one based on episodic long-term memory indices. Given these latent constructs, hearing impairment is related to a long-term memory deficit.

The results agree with a further prediction from the ELU model, i.e. that frequent mismatches would produce a relative disuse of episodic long-term memory. Hearing loss results in less precise phonological representations in long-term memory (Andersson (2002), and decreases the distinctiveness of the phonological input directly. Thereby, mismatch reduces the use and quality of the information encoded into, stored in, and retrieved from episodic long-term memory. In the long term, this may reduce the efficiency of episodic memory functioning.

Hearing loss does not seem to affect the recall of the items in short-term memory (Rönnberg et al., 2009b). This is consistent with the ELU model, because when mismatches occur, the general expectation is that explicit WM (or, more generally, short-term memory) processing will be increasingly important, and will hence not degrade due to disuse. Note that these conclusions are made cautiously, as the data are cross-sectional and preliminary. We intend to study the effect of the hearing aids per se in a forthcoming study, when we are able to match users with non-users on impairment and other relevant background variables.

Age and ELU

The general thesis is that both the implicit components of the ELU model, such as lexical access and temporal signal processing aspects (e.g. Salthouse, 1996; Pichora-Fuller, this volume), and the more explicit components such as WM capacity (e.g. Luo & Craik, 2008), are negatively affected by chronological age (Rönnberg, 1990). This is also borne out by evidence from another language modality, which shows that there is an age-related difference specific to deaf signers in the ability to retain order information in WM when temporal processing demands are high (Rudner et al., 2009a).

However, the interesting points that can be made from the ELU perspective are that (a) performance in
mismatch conditions does not in a relative sense rely on age or hearing thresholds; here, almost all variance is accounted for by WM capacity as indexed by reading span performance, whereas for match conditions, age and hearing thresholds are relatively more important (Lunner & Sundewall-Thorén, 2007; Rudner et al., 2009a,b, or c), (b) the effect of disuse of episodic long-term memory due to hearing loss is not confounded by age (Rönnberg et al., 2009b). Thus, phonological mismatch and the related disuse of phonological long-term memory seem to be important age-invariant determinants of performance, and are phenomena that are relevant to take into account with hearing aid fitting. Knowledge of these effects may help the clinician to disentangle the cognitive effects that may be modified as a function of age, and those that may not.

WM Training in Old Age?

One cognitive function that may be expected to be modified by training is WM capacity and related executive functions. Recent research on WM training shows behavioral as well as neural activity changes after 5–6 weeks of intensive training (Olesen et al., 2004), with concomitant increases in cortical dopamine receptor D1 binding (McNab et al., 2009) for persons without disability, and using the same computerized training programme also for patients suffering from acquired brain injury (Lundqvist et al., 2009). In that latter study, interesting transfer was obtained from visuospatial WM exercises to non-trained tasks such as listening span. This transfer may prove to be a crucial link that in future studies will be addressed, with the underlying hypothesis that it is possible to train abilities crucial to listening/reading span performance, which in turn is supposed to mediate improved speech understanding in old age. Cognitive hearing science research has just begun to explore these very challenging possibilities.

References


Pichora Fuller, M.K. (this book). Using the brain when the ears are challenged helps healthy older listeners compensate and preserve communication function.


Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue:
A framework and a model. *International Journal of Audiology*, 42; S68–S76.