Frontiers in Psychology

TYPE Review
PUBLISHED 01 September 2022
DOI 10.3389/fpsyg.2022.967260

Check for updates

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Rebecca E. Millman,
The University of Manchester,
United Kingdom
Arthur Wingfield,
Brandeis University, United States

*CORRESPONDENCE
Jerker Rönnberg
jerker.ronnberg@liu.se

SPECIALTY SECTION
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

RECEIVED 12 June 2022
ACCEPTED 08 August 2022
PUBLISHED 01 September 2022

CITATION
Rönnberg J, Signoret C, Andin J and Holmer E (2022) The cognitive hearing science perspective on perceiving, understanding, and remembering language: The ELU model.
Front. Psychol. 13:967260.
doi: 10.3389/fpsyg.2022.967260

The cognitive hearing science perspective on perceiving, understanding, and remembering language: The ELU model

Jerker Rönnberg*, Carine Signoret, Josefine Andin and Emil Holmer

Linnaeus Centre HEAD, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

The review gives an introductory description of the successive development of data patterns based on comparisons between hearing-impaired and normal hearing participants' speech understanding skills, later prompting the formulation of the Ease of Language Understanding (ELU) model. The model builds on the interaction between an input buffer (RAMBPHO, Rapid Automatic Multimodal Binding of PHOnology) and three memory systems: working memory (WM), semantic long-term memory (SLTM), and episodic long-term memory (ELTM). RAMBPHO input may either match or mismatch multimodal SLTM representations. Given a match, lexical access is accomplished rapidly and implicitly within approximately 100–400 ms. Given a mismatch, the prediction is that WM is engaged explicitly to repair the meaning of the input – in interaction with SLTM and ELTM – taking seconds rather than milliseconds. The multimodal and multilevel nature of representations held in WM and LTM are at the center of the review, being integral parts of the prediction and postdiction components of language understanding. Finally, some hypotheses based on a selective use-disuse of memory systems mechanism are described in relation to mild cognitive impairment and dementia. Alternative speech perception and WM models are evaluated, and recent developments and generalisations, ELU model tests, and boundaries are discussed.

KEYWORDS

the ELU model, working memory, semantic long-term memory, episodic long-term memory, adverse listening conditions, age-related hearing loss, dementia

Background

Cognitive hearing science builds on the principle that individual cognitive functions play an important role from very early subcortical auditory processing (Stenfelt and Rönnberg, 2009; Sörqvist et al., 2012) to interactions among memory systems at cortical levels of listening, language understanding, and dialogue (Rudner et al., 2008, 2009; Rönnberg et al., 2021, 2022). Anatomically, several precise downstream corticofugal pyramidal cell axons from neocortical layers 5 and 6 (Usrey and Sherman, 2019) allow for early cognitive impact at subcortical levels, even down to the cochlea (cf. the early filter model, Marsh and Campbell, 2016). This neural organization sets the stage for deep cognitive penetration of the very early sensory and perceptual windows of our experience—a possibility that had not been systematically scrutinized in the audiological and hearing research field before the advent of the Ease of Language Understanding model (ELU, Rönnberg, 2003). Generally, the ELU model (Rönnberg et al., 2008, 2013, 2019) is about rapid abstraction of the meaning of multimodal linguistic input, mediated by working memory (WM) in adverse listening conditions (Mattys et al., 2012).

Our early studies of speech perception and speech understanding focused on speech-reading, or lip-reading (a narrower term not including gesture and body language), adult individuals with normal hearing, impaired hearing, and deafness. The question that we posed was whether persons with hearing impairment—through their increased reliance on visual speech–produced superior, compensatory, visual speech perception/understanding skills. We also investigated the presentation modality, type of materials, type of task (Hygge et al., 1992; even including more ecological tasks Rönnberg et al., 1983), and whether hearing-impairment related variables like the duration of impairment and/or the degree of hearing loss played a role (e.g., Rönnberg et al., 1982, 1983; Lyxell and Rönnberg, 1987; Rönnberg, 1990). The answer was surprising since no compensatory signs were empirically observed.

The emergence of cognitive hearing science

With further data collections, the data pattern took a radically different turn: First, we tried to examine why some people were such excellent speech-readers (e.g., Rönnberg, 1993; Rönnberg et al., 1999). In a set of case studies of extreme speech-reading skills, we demonstrated that instead of a compensatory effect due to the hearing impairment, it was about cognitive skill in processing and storage of perceived information, measured by the reading span test (RST, Daneman and Carpenter, 1980; Daneman and Merikle, 1996). High RST performance described the cases who in their daily life relied on poorly conveyed auditory speech information, but who still were very competent communicatively: it could be lip-reading only (the case of SJ: Lyxell, 1994), tactile-visually conveyed speech information (the case of GS: Rönnberg, 1993), or a hearing-impaired person with a speech-sign bilingual background (the case of MM: Rönnberg et al., 1999). They all used very different communication strategies, but effectively so. The common denominator of the different case studies reported was that each person was well-equipped cognitively, and that cognitive functions seemed to operate over and above the variables we had studied up to that point. More specifically, it was demonstrated and replicated from the case studies that not only did high WM capacity (WMC) play a significant role in holding information alive, thus presumably mitigating the prediction of upcoming events, but it also represented a cognitive workbench for reconstructing misperceived linguistic units (i.e., postdiction, Rönnberg et al., 2019, 2021, 2022). In the same vein, we found that other related kinds of cognitive functions also contributed to the picture.

It was observed that cognitive functions like lexical access speed (Rönnberg, 1990), executive functions (Andersson and Lidestam, 2005), and inference-making capacity (Lyxell and Rönnberg, 1987) were associated with speech perception and understanding (reviewed in Rönnberg et al., 1998, 2021; Rönnberg, 2003). The data pattern withstood many experimental variations, especially in difficult speech-in-noise conditions (reviewed by Lyxell et al., 1996; Gatehouse et al., 2003, 2006; Akeroyd, 2008; Arlinger et al., 2009; Lunner et al., 2009; Besser et al., 2013), where WMC played the dominating role. Thus, it was (and still is, see e.g., Mishra et al., 2021) hard to escape the general conclusion that poor hearing and/or poorly specified or fragmented speech stimuli depend on individual cognitive processing skills to fill in the gaps of incomplete input to the perceptual and cognitive systems. These findings make up the foundation of Cognitive Hearing Science. For a more complete and historical account of the emergence of the field, see Arlinger et al. (2009).

Early studies of cross-modal language plasticity

Early neurophysiological evidence spoke to the issue of plasticity of brain tissue and prerequisites for commonalities in central perceptual and cognitive functions. Many studies testified to cross-modal language activations, suggesting for example that visual areas are recruited in pre-lingually deaf cochlear implant users (Giraud et al., 2000, 2001; Zatorre, 2001; Kral and Sharma, 2012). In addition, tactile stimuli in the congenitally deaf tactile aid user activate secondary auditory areas (Levänen, 1998). Also, the duration of deprivation plays a key role in the reorganization of the sensory cortices, such as early sensory deprivation will result in better neural plasticity adaptation (Tillberg et al., 1996; Bernstein et al., 1998;

MacSweeney et al., 2001). In normal-hearing listeners, it was suggested that silent lip-reading activates the auditory cortex (Calvert et al., 1997), especially for skilled speech-readers (Ludman et al., 2000). Furthermore, signed and auditory to-be recalled story materials perceived by sign-speech bilinguals (i.e., sign-language interpreters) have been shown to activate temporal areas to a similar extent if a visual component was involved for both modalities (Söderfeldt et al., 1994). However, compared to auditory-only, specific bilateral temporal areas activated by sign-language were involved, specifically the addition of the left area V5, later replicated across imaging techniques (e.g., Söderfeldt et al., 1997; Rudner et al., 2007, but see further under *boundaries*). However, when it came to WM for sign and speech, we still found that there were similarities for left inferior frontal and inferior temporal gyri, which subserve phonological and semantic processing areas (Rönnberg et al., 2004)—areas that also were similarly activated in the early Söderfeldt et al. (1994, 1997) studies. Finally, sign language phonological awareness and word reading ability have also been demonstrated to be associated (Holmer et al., 2016). Again, these data suggest that the brain rapidly transcends the "raw" sensory codes and rapidly abstracts input into modality compatible representations.

In all, there were many early studies that suggested commonalities and plasticity in brain activation independent of language modality and presentation modality. In addition, individual cognitive factors like WMC determined performance on language perception, and the WM system seemed to have modality-independent properties. These neurophysiological data patterns—in combination with the behavioral data—prompted the formulation of a modality-independent ELU model based on individual differences in specific perceptual and cognitive components (Rönnberg, 2003; Rönnberg et al., 2008).

The ELU model takes shape

In the formulation of the original ELU model (Rönnberg, 2003; Rönnberg et al., 2008), we were quite bold in the sense that the assumption of an occurring mismatch between perceived language input and long-term memory representations of linguistic units was supposed to hold across sensory modalities (auditory, visual, and tactile) as well as language modality (spoken and signed). The "language processor" in the brain was assumed to have a multimodal combinatorial capacity, typically occurring at the "syllabic" or sublexical level across language and presentation modes (Rönnberg, 2003; Stenfelt and Rönnberg, 2009). All cases of mismatch were supposed to trigger an increased dependence on WMC for reconstructive, postdictive purposes.

In some more detail, the ELU system assumes that the perceptual input is conceptualized as an input buffer which Rapidly, Automatically, Multimodally Binds PHOnological information together (RAMBPHO, cf. Baddeley, 2000, 2012; Stenfelt and Rönnberg, 2009). This binding, or integration process, presupposes a rapid "default mode" of abstraction into a multimodal input, where the main task of the system is to implicitly and directly unlock multi-attribute phonological representations in Semantic Long-Term Memory (SLTM), leading to access of lexical meaning (Bernstein et al., 1998; Bavelier and Neville, 2002; Stenfelt and Rönnberg, 2009; Rönnberg et al., 2013, 2019). This process typically occurs during a short time window from 100 to 400 ms, depending on the paradigm, if the chain of events runs smoothly, implicitly, and without effort. In general, this RAMBPHO process is reminiscent of Gibson's (1966) direct perception approach in that the senses should be considered as interacting perceptual systems, without short-lived intermediary representations.

However, for hearing-impaired participants, or when listening conditions are adverse (e.g., when competing noises or foreign accents are present, or the signal processing in the hearing aid is suboptimal), RAMBPHO-delivered attributes may be fuzzy and too few in numbers to surpass a hypothetical threshold to unlock lexical representations in SLTM (see Rönnberg et al., 2013, for details). The consequence of such a mismatch is that more deliberate, explicit and WM-based storage and processing functions are assumed to be triggered. These WM functions purportedly aim to piece together and infer what was communicated (Lunner et al., 2009; Rönnberg et al., 2013, 2019, 2021). These explicit functions may depend on several inference-making operations within WM but also include several interactions with SLTM and Episodic Long-Term Memory (ELTM), hence taking a relatively longer time than effortless implicit processing. The implicit processes typically operate on a millisecond scale, and the explicit processes may take seconds (Stenfelt and Rönnberg, 2009), and recent evidence suggests that different brain oscillations can dissociate the two (e.g., Gray et al., 2022). There will always be a ratio between the two, which is assumed to vary dynamically from moment to moment due to turn-taking and interlocutor responses in a conversation (Rönnberg et al., 2019).

In general terms, prediction (and postdiction) processes affect the probability that RAMBPHO will match or mismatch with SLTM representations. On a general time scale, RAMBPHO-delivered information always precedes and then affects WM storage and processing operations. If a mismatch occurs, slower postdiction processes in WM feed back to RAMBPHO until comprehension is reached (see more under theoretical implications) or not, for example in cases where the listener is not sufficiently motivated to allocate resources required for further speech processing (Pichora-Fuller et al., 2016). Thus, it is important to acknowledge that RAMBPHO is an obligatory part of an ELU/WM system, feeding linguistic information to the match/mismatch mechanism—which is at the heart of the system. The ELU model describes a communication system which relies on interacting memory systems and mechanisms.

Experimental evidence

The first experimental manipulations of habitual vs. non-habitual signal processing in hearing aids successfully tested the cognitive consequence of the mismatch notion. We developed several different kinds of methods to trigger a mismatch. One example, and probably the most evident demonstration, was that of the studies by Rudner et al. (2008, 2009). For example, experimental acclimatization to a non-habitual kind of aggressive hearing-aid signal processing for 9 weeks (i.e., FAST or SLOW wide dynamic range compression) and then subsequent testing in a previously non-acclimatized mode of signal processing (i.e., SLOW-FAST or FAST-SLOW), produced strong reliance on WM in those two mismatching conditions, compared to the matching FAST-FAST and SLOW-SLOW conditions. As a matter of fact, the effect of just shifting from the regular hearing-aid settings to new ones produced higher reliance on WMC (for replications and relevant supporting studies/reviews, see Souza and Sirow, 2014; Souza and Arehart, 2015; Rönnberg et al., 2019; Souza et al., 2019). These findings also expose the interplay between WM and SLTM in the development of representations, as introduced to the ELU framework in recent years (Holmer et al., 2016; Holmer and Rudner, 2020; Rönnberg et al., 2022). In response to a novel input signal, such as that produced by new settings in a hearing aid or learning a new word, the language system is likely to treat the input as something unfamiliar, i.e., a mismatch condition, which WM resources are used to solve. However, each time a mismatch condition is resolved, this has the potential of producing an adjustment to the exemplars associated with the representational space in SLTM.

A second example is the investigation of competing effects of different kinds of maskers. Many studies agree with our view that so-called energetic maskers (Brungart, 2001) produce distraction but not to the same extent as informational maskers engaging SLTM (e.g., Rönnberg et al., 2010; Mattys et al., 2012; Sörqvist and Rönnberg, 2012; Kilman et al., 2014), and that distraction is more pronounced if the masker was in the participants' native language (Kilman et al., 2014; Ng et al., 2015). It should be noted that the original data of WM dependence (using "speech-like" maskers) had already been observed and discussed (Lunner, 2003; Lunner and Sundewall-Thorén, 2007; see a review by Rönnberg et al., 2010). Again, in retrospect, the effects of informational or speech-like maskers were related to partial activation of SLTM, e.g., phonologically similar neighbors (Luce and Pisoni, 1998) or possibly, with ELTM repetitions of SLTM contents.

Thirdly, as noted above, WMC is also an important predictor of performance in ELTM in such circumstances of initial speech-in-speech maskers (e.g., four-talker babble, 4T, Sörqvist and Rönnberg, 2012; Ng and Rönnberg, 2020). Notably, the robustness of the high dependence on WM was not influenced by the duration of hearing-aid use, at least up to 10 years of hearing-aid use for four-talker (4T) maskers (Ng and Rönnberg, 2020). This raises the question of whether some kinds of conversational environments are too dynamic to allow for a lessened dependence on WM. That is, highly dynamic input—or input with poorly defined phonological information—might stress the boundaries of how well the system can adjust its representational space (Han et al., 2019). The 4T-masker results hold irrespective of signal processing in the hearing aid. However, contextual support (Rönnberg et al., 2016) or plausibility/predictability of sentences may override the need for WM resources (Moradi et al., 2013; Amichetti et al., 2016), which in turn makes the signal-context interactions determine the potential need for postdictive processing. The overall idea is that the brain should not invoke WM resources unnecessarily, e.g., when context drives a more rapid and implicit route to comprehension. In that sense, the brain is "lazy" and economical in spending effort and processing energy, using a principle of least effort (cf. Ayasse et al., 2021; Silvestrini et al., 2022).

A final study points to constraints when bimodally combining CI-listening in one ear with listening with a hearing aid in the other (Hua et al., 2017). These two types of signals reaching the brain will not necessarily be RAMBPHO compatible. From an ELU perspective, an electric and a physical-neurostimulation might be harder to convert into some more abstract representation than naturally occurring multimodal sensory stimuli. Analogous to habitual vs. non-habitual signal processing in the hearing aid, this combination of inherently different signals to the brain does not seem to combine easily. One indication is that RST was the most sensitive predictor variable for bimodal sentence materials compared to unimodal listening conditions, where the trail-making test (primarily measuring cognitive speed) was more critical to unimodal conditions and single word identification (Hua et al., 2017). But, it should also be noted that the bimodal condition facilitated speech-in-noise performance, although the cost in terms of WM engagement and effort may create a balancing act that the individual and clinician must decide from the individual WMC data.

Related to these difficult (or mismatching) speech processing tasks is a couple of recent studies corroborating the ELU hypothesis about the engagement of WM in challenging listening conditions. Mishra et al. (2021) found that WMC accounted for large portions of variance (up to 80%) of word recognition in speech noise of spondees and phonetically balanced word lists presented at dB SNR 0, −10, and −20, whereas in quiet the pure tone average explained 78% of the word recognition scores and not the RST scores (see also Kurthen et al., 2020). However, there is also evidence for specificity in correlations between the demands of the WM task and the complexity of the criterion task (Heinrich et al., 2015; see Rönnberg et al., 2021 for a discussion). Thus, the actual principle of mismatching is replicated in Mishra et al. (2021) with the concomitant demand on WM resources, but further task analyses of both tests of WMC and the type of outcome remain to be carried out (Heinrich et al., 2015). Nevertheless, see our first attempts involving interactions among memory systems (Rönnberg et al., 2011, 2014, 2021).

Linguistic abstraction and WM capacity

In gating tasks (Grosjean, 1980), the participants are required to identify a consonant/vowel or a final word in a sentence, based on the presentation of successive bits of initial phonetic information of the speech token (Moradi et al., 2013, 2014a,b). These studies have demonstrated that early and successful linguistic identification requires WMC. WM is involved in the rapid identification of e.g., a consonant when the semantic context is lacking. Around 90 msec is necessary to identify speech tokens in the auditory gating paradigm (Moradi et al., 2014b). Audiovisual presentation reduces this identification time to 40–50 msec, as does identification of final words in highly predictable sentences (Moradi et al., 2013). This time reduction presumably implies very rapid neuronal communication between different brain regions that, in turn, activate different memory systems. However, hearing loss (even when compensated with hearing aids), age, and noisy signals, all slow down the identification process (Moradi et al., 2013, 2014a).

A further example of rapid abstraction is a kind of priming paradigm which has demonstrated the so-called "pop-out" effect (see a general discussion in Davis et al., 2005), where presenting the written version of a sentence on a computer screen creates an enhanced perceived clarity and understanding of spoken, noise-vocoded, sentences that are otherwise incomprehensible due to the vocoding. In the studies by Signoret et al. (2018), Signoret and Rudner (2019), the written version of the sentence was presented word by word (200 msec before the vocoded version) until a sentence was complete, and required the participant to rate the perceptual clarity of the vocoded sentence. Perceptual clarity is enhanced for semantically coherent and phonologically primed sentences, but when compatibility is low, WM processes entered into play. For example, WM was invoked in conditions with non-matching primes, substantial vocoding, and low semantic coherence (Signoret and Rudner, 2019).

In a recent study using magnetoencephalography (MEG, Signoret et al., 2020), participants were to decide whether the final word in a sentence was expected or not. The participants had studied the sentences before the actual experiment, so expectations on the final word were strong. Different kinds of deviants were used as final words (in background noise): either the final word was semantically different but rhymed with the expected word, or was semantically related but did not rhyme, or was different in both aspects of similarity. Notably, WMC negatively correlated with the number of false alarms to meaning deviants that rhymed, such that participants with high WMC were less lured into accepting a deviant via the RAMBPHO to the SLTM matching process. Further, participants with higher WMC had processed meaning deviants more easily (smaller N400 effect) compared to participants with lower WM capacity. Participants with high WMC, also processed the semantic mismatching more easily (smaller N400 effects) and showed better performance at the behavioral level. WM therefore seems part and parcel of the prediction mechanism.

In sum, the hypothesis of RAMBPHO-like multimodal representations that together with high WMC activate SLTM, receive support from the above examples.

Theoretical implications

Theoretically and neurophysiologically, the RAMBPHO-SLTM interaction necessarily utilizes very rapid subcortical and cortical connections that allow for the implicit initial matching process. At this stage, the *prediction* aspect of the ELU model allow for matching processes most likely affecting attention to certain aspects of the input signal, which naturally varies in form, content, and modality (Samuelsson and Rönnberg, 1993; Sörqvist et al., 2012; Sörqvist and Rönnberg, 2012). Because of the correlations with WMC cited above, we can assume that RAMBPHO rapidly "constructs" a multimodal channel to WM that matches multi-attribute or multimodal representations in SLTM. Thus, a high WMC may facilitate early attention fine-tuning of auditory processing but may also reflect a highly synchronized brain network (Fell and Axmacher, 2011). The conclusion about some kind of early fine-tuning is reinforced by the finding that WM processes are positively interconnected with the effects of practice on auditory music skills (Kraus and Chandrasekaran, 2010) and their corresponding neural brain stem signatures (Kraus et al., 2012).

Thus, we argue that the brain always has, as its primary aim, to abstract meaning, i.e., a cognitive hearing, sense-making, organism. This aim holds regardless of whether the stimuli are represented by sublexical or lexical items, or grammatical constraints in a sentence (Ayasse and Wingfield, 2020). And, if there is a mismatch, there is the advantage of already existing multimodal representations maintained in WM—given a sufficiently capacious system—which can be deconstructed and/or reconstructed among brain networks that belong to SLTM and ELTM (i.e., postdiction).

Furthermore, since we initially have a presumed representation which is rapidly constructed but which also can be deconstructed by cues, and then reconstructed again, perhaps several times, it necessarily takes longer to execute explicit interactions among brain networks of the brain that belong to the WM, SLTM, and ELTM systems. The explicit postdiction process may take seconds, while the prediction process is on the msec scale. The neural mechanism diverges between them: while the postdiction process is proposed to relate to theta-activity, the prediction process is proposed to relate to alpha-activity (Gray et al., 2022). Typically, the postdiction process becomes more explicit and slower the neural process (i.e., probably related to enhanced theta activity), whereas the prediction processes are more implicit and probably related to faster neural activity, such as decreased alpha activity (Gray et al., 2022) or beta activity (Signoret et al., 2013). Although predictions are often primed implicitly in everyday conversation in their effect on RAMBPHO, contextual cues can be explicitly held in WM prior to the experimental materials (Zekveld et al., 2013), in a similar vein to when postdiction feeds back into RAMBPHO. This priming mechanism likely aims to pre-activate specific knowledge at different levels of processing depending on the environmental context and on individual abilities and skills. Generally speaking, the ELU model assumes an overarching prediction-postdiction system, both dependent on WMC, albeit in different ways (Rönnberg et al., 2019, 2022).

Output from the ELU system

The output of the ELU system is lexical access, the grasp of what was communicated, and what consequently may be recalled from ELTM. Further, the output might involve a change to SLTM, either during development (Holmer et al., 2016) or when adjusting to novel listening conditions (e.g., Ng and Rönnberg, 2020). We have also shown that WMC contributes to ELTM performance in terms of sentence recognition (e.g., Sörqvist and Rönnberg, 2012; Zekveld et al., 2013). In Sörqvist and Rönnberg (2012), it was demonstrated that performance on the Size Comparison (*SIC*) span WM test predicted higher immediate and delayed recall of fictitious stories masked by a person reading from another story at encoding—*SIC* span made significantly better predictions in hierarchical regression analyses the RST. This presumably comes from the fact that apart from processing and storage, as in the RST, *SIC* span involves an additional inhibition component. For example, the task could be to compare four-footed animals in a list of comparisons and an additional to-be-remembered word for each comparison: Is a zebra larger than a mouse? (the Comparison) + the word (lion). Each list of comparisons and words belong to the same semantic category, which could cause confusion between comparison words and to-be-remembered words at the recall of the list. Obviously, an inhibition factor comes into play, and mediates better recall.

Although not a prime purpose of that study (Sörqvist and Rönnberg, 2012), the clinical implications are important as well. As it happened in this study, WMC was important for both immediate performance and for ELTM. The ELTM aspect is crucial from a listening perspective. If the actual signal processing in the hearing aid or the adversity of the listening situation drains too much processing capacity in the "here and now" situation, then less is left over for storage. This implies that conditions for change might also be circumscribed in noisy conditions, perhaps because little WM resources are left for successful encoding into change and development of SLTM. To compare with a long discussion about task demands on storage and processing, see Lunner et al. (2009) and Rönnberg et al. (2021).

In other words, in the dialogue between a hearing-impaired person and an interlocutor, the hearing-impaired person must allocate explicit attention to mismatches when they occur to be able to extract meaning, and perhaps learn something new, from the conversation. This is not required of the normal hearing person to the same extent. This is exactly the reason why a measure of what is left in ELTM after smaller amounts of storage resources remain in WM (e.g., tested a day or two after the conversation) would clinically be a very important ecological aspect of what it means to approach ease of listening and understanding for a hearing-impaired person. Without going into detail, a test that tapped into storage, semantic processing functions, and inhibition at the same time, would seem to be a suitable candidate based on the data we have collected (e.g., Sörqvist and Rönnberg, 2012; Stenbäck et al., 2015). Furthermore, what you remember from a conversation has obvious personal and social consequences, and such consequences might sometimes lead to developing depression (Keidser et al., 2015).

In further support of the inhibition component, recent studies by Stenbäck et al. (2015, 2021) verify that especially WMC (measured with the RST) but also the Swedish Hayling test (Stenbäck et al., 2015), which measures inhibition, were significant predictors of performance in speech-in-noise (SPIN, Hällgren et al., 2006) and Hagerman matrix sentences (Hagerman, 1982). The Hayling test builds on the ability to inhibit sentence completion of the last semantically correct word instead of providing a semantically incorrect but grammatically correct word. Thus, WMC and executive functions are part and parcel of listening, understanding, and recalling in adverse conditions.

In general, inhibition and turn-taking in real dyads or conversation/discussion groups put large demands on the timing of turns. If you are, e.g., interrupting too many times, it might just be the case that you do not have sufficient WMC to follow the line of thought in the conversation. To do that, you need to keep in mind what was just said and process it, at the same time as you are planning to latch on with your own turn, and what you are going to respond to. Therefore, taking another person's perspective in a dialogue demands WMC functions of maintenance, timing and dual storage and phonological/semantic processing. However, taking the perspective of someone else also involves the cognitive function of theory-of-mind (ToM), which is about decoding and understanding other people's intentions and feelings, and not necessarily just grasping what was actually said in the conversation (Hagoort and Levinson, 2014). That is, the intended meaning might sometimes not be coded in the meaning of the exact wording of a sentence, and therefore, ease of language understanding is about multilevel understanding in dialogue.

Recent findings

Füllgrabe et al. (2014) and Füllgrabe and Rosen (2016) claimed that WMC only accounted for significant amounts of variance in elderly hearing-impaired participants' SPIN performance, whereas for younger normal-hearing participants, only a small percent of the variance was accounted for by WMC. Nevertheless, Vermeire et al. (2019) clearly showed that for elderly normal-hearing participants, RST was a significant predictor of SPIN performance as well. Indeed, Gordon-Salant and Cole (2016) showed the same results to hold across age groups, with RST as part of the most prominent predictors. In the same vein, with large samples, Marsja et al. (2022) used the n200 database (Rönnberg et al., 2016) to study potential differences in cognitive involvement due to hearing loss. Marsja et al. (2022) used a multi-group structural equation model (SEM) approach where the purpose was to assess whether the contribution of a "Cognition" latent variable (based on RST, a visuospatial WM test and a semantic WM word-pair test, and Raven's matrices) was equally related to a SPIN criterion (Hagerman matrix sentences, Hagerman, 1982) for hearing-impaired hearing-aid users compared to normal-hearing participants. The results, based on 200 participants per group, show that the Cognition variable accounted for identical beta weights (-0.32) in both groups of equal average age (60 years), when the groups were compared on an outcome latent construct based on Hagerman matrix sentences. Thus, the cognitive contribution to SPIN perception is not specific to elderly hearing-impaired participants. The statistical models were partialed out for age and hearing loss, and significant on all relevant model fit parameters. This is generally supportive of the initial claims of the ELU model, viz. that there is a communality in cognitive abstraction and cognitive prediction across adult groups with different hearing status (Rönnberg, 2003).

The devil is in the details

It is important to note that the interaction between the fine details of task demands may make a large difference in terms of predictability of outcome in a SPIN task. For example, in the original RST (Daneman and Carpenter, 1980), participants always recalled the final words in each sentence set—which obviously invites a strategic component compared to the version we use (Rönnberg et al., 2016), where participants are post-cued to recalling *either* the first *or* the last word of the sentences to be verified. In the latter case where the strategic component is reduced, the "raw" WMC is more likely to be revealed. Our research builds on this latter task version and that "detail" may be a clue as to why some researchers get higher involvement of WM than in some other studies (e.g., Ng et al., 2013; Souza et al., 2019; Ng and Rönnberg, 2020). Other aspects relate to contextual support either at the prediction stage or in the sentence materials themselves; high contextual support renders lower correlations with WMC, and vice versa (Moradi et al., 2013; Rönnberg et al., 2016). Dependence also varies with age, hearing status, and a host of other factors related to hearing aid signal processing and habitual processing demands, and not least the interplay amongst the speed, phonology, and WM factors depends on the level of adversity of the listening situation (Homman et al., submitted).

Structural equation modeling

In a recent study by Homman et al. (submitted) on the hearing-impaired participants in the n200 study (Rönnberg et al., 2016), we used Structural Equation Modeling (SEM) based on the original cognitive parameters in the ELU model: speed, phonology, and WM (Rönnberg, 2003). Thus, one latent speed parameter (i.e., physical matching and lexical access speed), one latent phonological parameter (auditory and audiovisual gating conditions and rhyme tests, i.e., measures of RAMBPHO), and one latent WMC parameter (i.e., based on the RST, visuospatial WM, and semantic word-pairs), were included, while age and hearing loss were partialed out. The results show that phonology always contributed to the performance in the different Hagerman conditions (irrespective of noise type, performance level, and type of signal processing). Speed did not directly predict the Hagerman outcome, but speed always predicted WM, and the WM to Hagerman path was significant only in the more difficult listening conditions involving 4T maskers. Thus, the new and interesting result of this mediation analysis was that speed contributed *via WMC* to Hagerman in the difficult conditions, i.e., where higher degrees of mismatch can be assumed. The general interpretation is that when being exposed to adverse listening conditions, it is important that WM is capacious because it takes more time to reconstruct what was perceived, i.e., when a more laborious explicit mode of processing is needed. An alternative interpretation would be that the adversity of the listening situation primarily strikes at RAMBPHO. Nevertheless, optimizing speed in WM operations becomes critical in both cases. This result agrees with previously reported results in Rönnberg et al. (2016), where WM was more strongly related to Hagerman matrix sentences than to HINT sentences, which are contextually driven everyday sentences. By virtue of the semantic coherence in HINT sentences, the prediction mechanism is improved, hence lessening the demand on WM resources for postdiction (cf. also Moradi et al., 2014b).

In a similar SEM approach, Janse and Andringa (2021) modeled word recognition performance in degraded low-pass filtered conditions. They used cognitive speed, vocabulary, hearing acuity, and WM as latent constructs. In their model, WM was the strongest latent construct relating to word recognition in noise, replicating our research. The RST was the test that loaded the highest on the WM factor, compared to digit span and non-word recall (cf. Rönnberg et al., 2016). In our current model (Rönnberg et al., 2021, 2022) we only used WM tests that emphasize storage and processing in dual task formats. We noted that speed of access from SLTM was our mediating factor. However, their mediation model of vocabulary via WM to word recognition (Janse and Andringa, 2021) is not directly comparable to ours, as we did not use vocabulary, but interesting indeed. The communality is that WM predictive capacity is only predictive of SPIN performance, via some back-up parameter such as SLTM speed or SLTM vocabulary. It is obvious that these mechanisms support WM when more complex interactions between WM, SLTM, and ELTM are required for postdiction purposes.

Comparison with other models

Perception and understanding: multimodal and multilevel aspects

Speech perception models are less comprehensive than the ELU model, but in some cases more specific. For example, in the Neighborhood activation model (NAM model, Luce and Pisoni, 1998), lexical access is clearly dependent on how input stimuli matches/mismatches with the lexicon due to phonological similarity and semantic parameters like word frequency. In the initial word cohort model (e.g., Marslen-Wilson, 1987), the initial information that enters the ear is assumed to activate a set of competitors in a cohort of possible candidates, a functional parallelism in the activation of the lexicon. As information successively enters the auditory system, activation and selection of candidates proceeds until only one lexical candidate remains. There are many manipulations of the selection process e.g., by priming, word length, or word endings that in different ways manipulate the word recognition point, which often occurs before the whole word has been perceived.

Furthermore, the probability of lexical access is not an all or none process (as described in Rönnberg et al., 2013); it depends on the RAMBPHO input and the kinds of representations it meets in LTM. And, at some hypothetical threshold of matching attributes lexical access is triggered. The types of error responses we obtain may well be captured by the NAM (Luce and Pisoni, 1998), but in addition we have also made clear that the prediction- postdiction cycle may prime or direct the individual ELU system to other aspects of representation in LTM that then helps the system to surpass the threshold and retrieve the correct lexical candidate. A good example of such priming by sentence context can be found in a MEG study by Signoret et al. (2020), where phonological and semantic error responses were in focus.

Also related to RAMBPHO, we focused on a special form of priming. We dubbed the hypothesis “perceptual doping” (Lidestam et al., 2014; Moradi et al., 2019). In brief, the priming effects of exposure to two initial conditions (auditory only, or audio-visually presented materials) on later auditory perception of consonants, vowels, and sentence materials generally demonstrated a multimodal facilitation (“doping”) effect. The interpretation is that there is a recalibration/remapping of the initial audiovisual presentation mode affecting the SLTM representation of phonological and lexical attributes. With the advantage of hindsight, a discussion of the data based on RAMBPHO may also have been possible.

Related to our mismatch concept, earlier basic auditory perception studies outlined the basic properties of the mismatch negativity (MMN) effect measured with EEG (Näätänen, 1995; Näätänen and Escera, 2000). In our research, we have emphasized the *consequence* of mismatch in terms WM involvement (Rönnberg, 2003). Relevant to the current paper is that the mismatch notion has also been applied to grammatical levels of language processing (Federmeier, 2007), as well as phonological/semantic processing (Signoret et al., 2020). The ELU model is here proposed to be about levels of linguistic mismatch, from RAMBPHO and lexical access, via grammar to semantic coherence. Therefore, the fact that the functional role of the frontal cortex in pre-attentive auditory change detection has been shown for grammatical deviations is of high importance (Hanna et al., 2014). The mismatch negativity function automatically detects grammatical anomalies around 200 msec, after the grammatical violation point (subject–verb agreement violations or word category violations, Hasting et al., 2007; cf. Signoret et al., 2020 for different violation types). Tse et al. (2013) and Hanna et al. (2014) have both demonstrated and discussed the very early, pre-attentive and automatic Broca/inferior frontal signals of mismatch negativity for grammatical violations (around 200 msec), which could be an inspiration to our new, more elaborated ELU proposal, of co-occurring mismatch signals possible at different linguistic levels (see below).

In our current view, the multimodal phonological level is crucial to SPIN performance, but if implicit processing occurs at higher levels of language like syntax, keeping auditory characteristics under control (Hasting et al., 2007), it makes our claim about the necessity of rapid WM interactions with SLTM and ELTM even more important. Otherwise, these extra steps would probably prolong the extra time for reconstruction and postdiction. This specification of the ELU model is that not only is WM involved in rapid RAMBPHO-delivered multimodal abstraction, but it is also involved in multilevel language interactions, given that there is some central mismatch time widow for several levels of language, and which can be processed in parallel (cf. Marslen-Wilson, 1987). *We submit that the prediction-RAMBPHO-SLTM-postdiction interaction demands a "moving time window" within the confines of WMC, the contents of which are rapidly abstracted at multimodal and multilevel aspects of input. In a more generalized form, it may be stated that: for any given aspect of RAMBPHO-delivered linguistic information, the cognitive consequence of a mismatch with SLTM representations, is bound to initiate WM-based postdiction.*

Furthermore, the cognitive consequence of a central multimodal/multilevel mismatch mechanism has not been fully realized in the ELU model, but comparisons with Central Auditory Processing Disorder (CAPD) research could inspire (Gates et al., 1996; Gates, 2012). However, several such multimodal and multilevel interactions will need to exploit all the storage and processing capacities of WM. Therefore, the postdiction processes will necessarily take more time than implicit predictions. But if there are rapid multilevel mismatch functional capacities of the brain, it will allow for an advanced analysis-by-synthesis kind of model, demanding such on-line revisions of what is misperceived (cf. Hickok and Poeppel, 2007). It also demands parallel processing not just at the word level. For example, in the Moradi et al. (2014b) study, high predictability sentences were completed with only minimal initial phonemic information of the final word (40 msec).

Thus, we still assume that relatively context-free perception is dependent on RAMBPHO-based lexical retrieval. But, context-bound, grammatically incorrect sentences can also induce mismatch at some violation point in the sentence. This implies that the functional parallelism (cf. Marslen-Wilson, 1987) is not only realized through multimodal streams of information, as in the ELU, but also in parallel streams at different levels of language that act in concert to optimize implicit understanding of the discourse. This may at later stages demand cognitive functions to keep track and focus on the "winning stream" of information processing (cf. Moradi et al., 2013). Again, presumably the brain is optimizing speed of mental operations in WM even in mismatch situations. However, to our knowledge, the mismatch studies have not emphasized the communicative feedback, which the ELU model denotes as postdiction, which in turn is assumed to feed back into predictive RAMBPHO processing. This postdiction feedback may not only alter predictions but also induce SLTM changes.

Thus, even when well-organized and linguistically interactive and smooth processing is taking place, mismatch at some linguistic levels will demand reconstruction and postdiction, typically taking more time. Parallel levels of processing (without any mismatch) may on the other hand synergistically integrate input and reduce processing time (Moradi et al., 2014a,b; Signoret et al., 2020).

Working memory

In comparison with other working memory models, the ELU model is very much inspired by two working memory traditions, the Baddeley and Hitch (1974; Baddeley, 2012) tradition and its many developments (e.g., the episodic buffer Baddeley, 2000, cf. RAMBPHO), as well as the tradition following a more general resource model tradition, with less structural assumptions on dedicated loops and modular functions (cf. our use of the RST, Daneman and Carpenter, 1980; Just and Carpenter, 1992; Daneman and Merikle, 1996; Barrouillet and Camos, 2020).

More specific capacity models sometimes have taken the form of activation of LTM relevant information, not seldom related to expertise (Ericsson and Kintsch, 1995; Cowan, 2005; Jones et al., 2007). The activation capacity of several representations in LTM thus becomes a measure of expertise, or WMC. In the ELU model, there are two roads to LTM in principle, one implicit and one explicit. This conceptualization and difference to the above models is dependent on the assumption that the ELU model is primarily conceived for communication purposes, where mismatch disturbs the flow of rapid phonologically mediated lexical access, but where WM must engage SLTM and ELTM to optimize explicit postdictions, as well as predictions. Seen from this horizon, the ELU model captures what we believe to be a human propensity, viz. the system is "lazy" or economical (cf. Richter, 2013); it does not spend explicit resources unless sub-threshold levels of language input cause mismatch (especially the phonologically mediated lexical access function).

There have been several recent attempts at refining the component concepts of the ELU-model. The ELU-model has generated several important scientific hypotheses and ways of investigating and testing them.

Model refinements

Edwards (2016) suggests that just before RAMBPHO processing occurs, a process is needed that accomplishes early perceptual segregation of the auditory object from the background, so called Auditory Scene Analysis (Dolležal et al., 2014). His discussion is based on the Rönnberg et al. (2008) version of the ELU model, where RAMBPHO processing focuses on how different streams of sensory information are integrated and bound into a phonological representation (see also Stenfelt and Rönnberg, 2009). Nevertheless, in Rönnberg et al. (2019, 2022), it is made more explicit that the system may feedback via postdiction processes, which may prime the prediction process (Sörqvist and Rönnberg, 2012), including fine-tuning of attention (Holmer and Rudner, 2020; Andin et al., 2021) and selection processes to specific features of the input (Rönnberg et al., 2013). This seems to be rather close to stream segregation, but the theoretical languages differ. By inference, postdiction may then calibrate the selection of the auditory object, comparable to "perceptual doping" (Moradi et al., 2019).

The second aspect is that RAMBPHO is assumed to be primarily dedicated to phonologically relevant information, embedded in lexical and semantic representations in SLTM. Lexical access and semantic meaning of sentences are tightly tied to the mismatch mechanism—and by default—finding of a linguistic object. Thus, that aspect of Edwards (2016) proposal does not necessarily demand model change (Rönnberg et al., 2019).

In the D-ELU model (Holmer et al., 2016), the development of language representations in SLTM is in focus. The original ELU model focused on the system's input side and the WM-LTM interactions, but the development of appropriate SLTM representations has hitherto received less interest. Nevertheless, it has been demonstrated that vocabulary is very important to speech perception in noise (Kennedy-Higgins et al., 2020), either via WMC (cf. Janse and Andringa, 2021), for hearing-impaired listeners (Signoret and Rudner, 2019), or in how language is represented in bilinguals (Kilman et al., 2014; Bsharat-Maalouf and Karawani, 2022). According to the D-ELU model, existing lexical representations in SLTM shapes further lexical growth, i.e., novel representations build upon existing representations (cf. Jones et al., 2021). Novel words that are rich in lexical attributes are more likely to be successfully encoded into SLTM, and thus learning rates are predicted to be steeper. Further, learning for persons with hearing loss is predicted to be worse than for controls when the perceptual platform at the learning stage is too dynamic (Ng and Rönnberg, 2020).

In the study by Kilman et al. (2014), and of relevance for how representations develop, we found in Swedish native speakers who also knew English, that the most interfering speech in noise condition was when the speech masker was in the same (Swedish) native language as the target. The Swedish babble was interfering more than the English babble in stationary noise, and in fluctuating noise. The interference from language maskers replicates previous work (Van Engen and Bradlow, 2007; Calandruccio et al., 2010).

A recent study by Bsharat-Maalouf and Karawani (2022) examined the speech perception of 60 Arabic-Hebrew bilinguals and a control group of native Hebrew speakers during degraded (speech in noise, vocoded speech) and quiet listening conditions. There was a clear interaction in the data such that performance in the bilinguals was on a par with the native Hebrew speakers in quiet conditions, whereas performance in the babble noise conditions (same language of the noise and targets) was substantially lower. Explaining these and other effects, in terms of proficiency (Kilman et al., 2014) of second language, age of acquisition, propensity to learn in vocoding conditions (Bsharat-Maalouf and Karawani, 2022), and what kinds of SLTM representations mediate these findings is of importance for the bilingual and developmental aspects of the ELU model. Future publications will tell.

Aging, cognitive impairment, and dementia

The ELU emphasis on a meaning-related focus of the brain's perceptual-cognitive system is assumed to prioritize multi-attribute representation and multilevel mismatch processing. In other words, both children and adults are primarily tuned in to understanding language and intended communication but can of course be instructed to learn or memorize what has been communicated. In terms of a use/disuse principle (Rönnberg et al., 2011, 2021), WM is on top, always dealing with both pre- and postdiction processes on-line; the next memory system is SLTM due to the natural semantic bias in interpretation of conversation and discourse, and ELTM will be relatively less used for two reasons: (1) a non-prioritized bias in communication, and (2) denied encoding and retrieval due to hearing loss or other adverse conditions. Thus, disuse can be a key to why WMC is relatively spared when it comes to cognitive decline studies (Rönnberg et al., 2011, 2014), whereas semantic and especially ELTM decline becomes a marker of mild cognitive impairment, which might develop into dementia.

The disuse notion of memory systems is mainly supported by two major studies by our team: (1) In Rönnberg et al. (2011), based on the Betula prospective cohort study (Nilsson et al., 1997), we found that hearing loss did not selectively affect different ELTM encoding tasks in different sensory modalities (i.e., motorically, by text and simultaneous auditory presentation, and auditory only compensated with hearing aids). If anything, the hearing loss–ELTM encoding task correlations were higher with the motorically encoded task. This may seem counterintuitive, unless one assumes multimodal representations and that the multimodal memory system level is negatively affected, not memory via a specific encoding modality. (2) In addition, long-term memory, especially ELTM was affected by hearing loss, but not by visual impairment. The fact that the cognitive aging and dementia-related literature suggests that ELTM is the most sensitive predictor variable among memory systems to mild cognitive impairment and dementia (Bäckman et al., 2001; Fortunato et al., 2016; Younan et al., 2020) makes our case strong. Combining (1) and (2), we may infer that hearing loss is an important risk factor for accelerated dementia progression (Livingston et al., 2017).

Common cause accounts (e.g., Baltes and Lindenberger, 1997; Humes, 2013; Powell et al., 2021) may predict that hearing loss affects several encoding modalities, but they do not predict selectivity of memory systems. In Rönnberg et al. (2014)—building on 138098 participants from the UK Biobank resource—we observed that ELTM was more affected by hearing loss than WM, thus replicating the data from Rönnberg et al. (2011). In terms of encoding modality, the Rönnberg et al. (2014) study employed visuospatial tests only. Still, we obtain a negative effect of hearing loss on ELTM and not on WM. This further supports and replicates a memory systems account of relative use/disuse as a potentially viable ELU explanation of the data pattern.

In addition, the potential risk of cognitive decline due to hearing loss cannot be explained by the information degradation hypotheses (Pichora-Fuller, 2003; McCoy et al., 2005), nor by attention costs for the hearing-impaired person (Sarampalis et al., 2009; Tun et al., 2009; Heinrich and Schneider, 2011). This could have been the case had the auditory encoding condition been negatively affected by hearing loss, even if the participants wore hearing-aids at testing (Rönnberg et al., 2011). A further important aspect of the 2011 data is that testing the same models, replacing hearing loss with estimated visual impairment (legibility of font size, on a scale from 6 to 24, Rönnberg et al., 2011; wearing eye glasses or not, or having a diagnosis, Rönnberg et al., 2014), did not replicate the memory system selectivity of the hearing loss results. As a matter of fact, the models tested were not acceptable by the structural equation model criteria used. What is also true of the above two data sets is that the hearing losses were only of the mild to moderate kinds (assessed by the pure tone audiogram in Rönnberg et al., 2011, and by the digit triplets test in Rönnberg et al., 2014), suggesting that early prevention with hearing aids should be employed (Arlinger, 2003), although the data for treatment by hearing aids is relatively meager when it comes to dementia.

At any rate, hearing loss, not visual impairment, is a very sensitive predictor variable of especially ELTM impairment, hearing loss being the largest modifiable factor of the development of dementia (Livingston et al., 2017). However, that is the overall picture and the more specific underlying mechanism as to why hearing loss is a risk factor for dementia is still argued to be unclear (Wayne and Johnsrude, 2015; Hewitt, 2017). Other independent analyses from the UK Biobank resource suggest that subclinical small variations in hearing acuity may still be associated with loss of gray matter volumes in the brain, especially in areas related to cognition and hearing (Rudner et al., 2019; however, see further about brain atrophy and cognitive reserve Uchida et al., 2021).

Generalizations

The previously mentioned study by Marsja et al. (2022) suggests impressively similar (if not identical) cognitive predictions from one hearing-impaired group compared to a normal-hearing group on a matrix sentence latent construct. This finding suggests a powerful generalization of the case that Cognitive Hearing Science—and the ELU model—applies to anyone, regardless of hearing status.

Moreover, recent studies of different speech distortions (Kennedy-Higgins et al., 2020) show that WM and vocabulary (i.e., SLTM) come out as the main predictors, irrespective of the type of distortion (time-compressed and noise-vocoded signals, and speech in noise). This informs us that the cognitive machinery underlying speech perception and speech understanding is rather invariant in its reliance on certain cognitive building blocks irrespective of how underspecified or distorted target stimuli are. As already argued, it presupposes that rapid abstraction into formats suitable for WM is a prerequisite for the system to work.

An interesting extension of the generalization aspect is the study by Blomberg et al. (2019) of adults with Attention Deficit Hyperactivity Disorder (ADHD). She also used different kinds of speech distortions (normal vs. noise-vocoded), orthogonally combined with type of background noise (clear speech, white noise, and speech babble). Materials were taken from the Swedish HINT sentence corpus, which consists of everyday sentences (Hällgren et al., 2006). Results showed that compared to an age-matched control group there was no interaction between group and type of masker or stimulus distortion (but main effects were observed), generalizing the Kennedy-Higgins et al. (2020) findings to another group of participants, with similar kinds of distortion manipulations. This pattern may depend on the possibility that the cognitive analysis and representations are multimodal and information-based rather than modality-specific. Importantly, different assessments of WM were used to construct a cognitive factor that heavily influenced performance across the distortion/noise conditions, supporting the ELU model.

As long as information collated or bound by RAMBPHO is incomplete in some of the many ways that will cause mismatch, dependence of WM tests indicate that a certain level of generalization is possible to make. But, we would not argue that RAMBPHO processing of e.g., vocoded speech is exactly the same as e.g., RAMBPHO processing of rapid wide dynamic range compression of speech. The general point is that at a cognitive postdiction level you must (for different reasons) infer, manipulate, and "mentally fill in" some pieces of information that demand WM processing, as well as retrieval of LTM information, to reconstruct poorly specified stimulus materials. Any other model that emphasizes the cognitive work needed in degraded, distorted, or perceptually demanding conditions would also be supported by such findings across groups and stimulus conditions (e.g., the FUEL framework, Pichora-Fuller et al., 2016).

Finally, another experiment seems to indicate that load on WM is reflected in larger pupil dilation responses (assumed to reflect cognitive load) than the physical characteristics (SNR) of the task (Zekveld et al., 2018), implying that high level cognitive processing in WM is accomplished, but pushes the system to its limits for participants with low WMC. This study was followed up by a study on an auditory Stroop task (measuring executive control), where pupil size was higher in conflict conditions (e.g., saying "left" in the right ear). This connects well with our early observations that not only did WMC play an important role in speech understanding, but also executive functions or cognitive control seemed important (Badre, 2021). In ELU terms, the postdictive phase of inferring what was uttered, may use both the processing capacities of WM but also of related or overlapping executive and cognitive control functions.

Testing the boundaries of the ELU model

Neuroimaging studies show that sensory deprivation, e.g., deafness, during development cause reorganization of superior temporal regions (auditory cortex; Bavelier and Neville, 2002; Merabet and Pascual-Leone, 2010; Andin and Holmer, 2022). Theories behind such cross-modal reorganization differs, with some suggesting pure neural processes and some suggesting behaviorally driven processes. In the case of early deafness, the latter has gained most attention, with two main lines of explanations (see extensive review in Cardin et al., 2020). The first explanation proposes functional preservation, where the type of processing in a sensory deprived region, i.e., auditory cortex, is preserved but applied to a different modality (e.g., visual instead of auditory). This notion finds support in results suggesting that superior temporal regions, which respond to speech in hearing individuals, are activated in response to sign language in deaf but not hearing signers (MacSweeney et al., 2001; Cardin et al., 2013). Such reorganization supports an extension of the ELU model to the manual-visual language modality.

The second proposal is that reorganization reflects a functional shift. This idea is supported by studies reporting activation in superior temporal regions during cognitive tasks (Twomey et al., 2017), and suggests modality-dependent differences in cognitive processes. This perspective speaks against one of the original claims of the ELU model, i.e., that there is a modality-independent "language processor" in the brain. This is of course given that superior temporal regions are exclusively engaged in language processing. However, the empirical evidence to date lends support for both explanations and it has also been suggested that they can coexist (Cardin et al., 2020).

In WM studies using sign-language material (e.g., Rönnberg et al., 2004; Bola et al., 2017; Cardin et al., 2018; Andin et al., 2021), the superior temporal regions (auditory cortex) and occipito-parietal regions (in speech-sign bilinguals, Rönnberg et al., 2004) are activated to a greater extent for deaf compared to hearing individuals. However, in a recent neuroimaging study from our lab, we found that the activation of auditory cortex did not increase with increasing WM load in a sign language-based task, suggesting a general sensory-perceptual processing role in response to visual linguistic material (Andin et al., 2021) in line with functional preservation. Further, we found support of a modality-specific pattern in relation to the degradation of the sign-language signal. In previous studies on auditory signal degradation in individuals with normal-hearing and impaired hearing, similar changes in neural activation have been identified for both increased WM load (amount of information needed to be kept in memory) and acoustic degradation. These findings have been taken as evidence for resource models of WM in general and the ELU model for language processing in particular (Obleser et al., 2012; Petersen et al., 2015; Peelle, 2018; Rönnberg et al., 2019). Although visual degradation of the language signal resulted in similar effects at the behavioral level, the neural overlap was absent for sign language in deaf early signers (Andin et al., 2021). Hence, while increasing WM load was reflected in increased engagement of the frontoparietal working memory network, as predicted, the degradation of the visual signal instead caused activation of bilateral inferior occipital and temporal cortices. The lack of neural overlap, might challenge the validity of the ELU model, potentially reflecting modality-specificity. However, it should be noted that the same effect was found for hearing non-signers. Hence, the effect might be related to presentation modality rather than the language modality. Further studies investigating the auditory and visual domain within the same paradigm are needed to further evaluate the modality-generality of the ELU model.

Conclusion

1. Cognitive and communicative data patterns preceding the formulation of the ELU model (Rönnberg, 2003) were described. Individual cognitive ability was (and is) important for communicative competence.
2. Rapid multimodal and multilevel abstraction by means of RAMBPHO is supported by recent and previous experiments. WM stores these types of information in an on-line "moving window."
3. Parallel levels of mismatch negativity make the system extremely effective and rapid in deconstruction and reconstruction, prediction and postdiction.
4. A use-disuse principle was introduced and combined with a multimodal memory systems account to suggest why hearing loss strikes at ELTM, SLTM, and WM in that order of decreasing negative impact.
5. Recent preliminary modelling gives strong and more nuanced support of a mediation model of the original ELU parameters, which takes into account that processing speed is important for WM operations only in adverse SPIN conditions Phonology (i.e., RAMBPHO) is a

basic predictor variable of SPIN performance under all circumstances.

6. New models, such as the D-ELU were discussed. SLTM adaptations show acclimatization to certain non-habitual signal processing strategies, as well as to "perceptual doping."
7. Language proficiency and bilingualism are further factors discussed in the D-ELU context.
8. Generalization studies have shown that hearing-impaired and normal hearing persons equally on a cognition factor as predictor of SPIN performance. Moreover, the reliance on WM across different signal distortion conditions is equal when comparing persons with ADHD and normal hearing persons.
9. Boundary conditions are discussed in a sign language context in terms of preserved brain functions which are applied to another language modality; or in terms of a functional shift, where deaf participants' sign language use is assumed to change brain organization.

Author contributions

JR conceived and wrote a primary draft of ELU-related research and model development. JA was especially responsible for the sign language research. CS was responsible for the ERP and MEG studies. EH was responsible for overall coherence and the D-ELU model. All authors contributed to the article and approved the submitted version.

Funding

Writing of this manuscript by JR was supported by the Swedish Research Council (Grant No. 2017-06092; held by Anders Fridberger).

Conflict of interest

Publisher's note

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47, S53–S71. doi: 10.1080/14992020802301142

Amichetti, N. M., White, A. G., and Wingfield, A. (2016). Multiple solutions to the same problem: Utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Fron. Psychol.* 7:789. doi: 10.3389/fpsyg.2016.00789

Andersson, U., and Lidestam, B. (2005). Bottom-up driven speechreading in a speechreading expert: The case of AA (JK023). *Ear Hear.* 26, 214–224. doi: 10.1097/00003446-200504000-00008

Andin, J., and Holmer, E. (2022). Reorganization of large-scale brain networks in deaf signing adults: The role of auditory cortex in functional reorganization following deafness. *Neuropsychologia* 166:108139. doi: 10.1016/j.neuropsychologia.2021.108139

Andin, J., Holmer, E., Schönström, K., and Rudner, M. (2021). Working memory for signs with poor visual resolution: FMRI evidence of reorganization of auditory cortex in deaf signers. *Cerebral Cortex* 31, 3165–3176. doi: 10.1093/cercor/bhaa400

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—A review. *Int. J. Audiol.* 42, S17–S20. doi: 10.3109/14992020309074639

Arlinger, S., Lunner, T., Lyxell, B., and Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scand. J. Psychol.* 50, 371–384. doi: 10.1111/j.1467-9450.2009.00753.x

Ayasse, N. D., Hodosn, A., and Wingfield, A. (2021). The principle of least effort and comprehension of spoken sentences by younger and older adults. *Front. Psychol. Lang. Sci.* 12:629464. doi: 10.3389/fpsyg.2021.629464

Ayasse, N. D., and Wingfield, A. (2020). The two sides of linguistic context: Eye-tracking as a measure of semantic competition in spoken word recognition among younger and older adults. *Front. Hum. Neurosci.* 14:132. doi: 10.3389/fnhum.2020.00132

Bäckman, L., Small, B. J., and Fratiglioni, L. (2001). Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain A J. Neurol.* 124, 96–102. doi: 10.1093/brain/124.1.96

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A., and Hitch, G. (1974). "Working memory," in *Psychology of learning and motivation*, 8, ed. G. H. Bower (Cambridge, MA: Academic Press), 47–89.

Badre, D. (2021). "Brain networks for cognitive control: Four unresolved questions," in *Intrusive thinking: From molecules to free will Strûngmann forum reports*, eds P. W. Kalivas and M. P. Paulus (Cambridge, MA: The MIT Press), 203–228. doi: 10.7551/mitpress/13875.001.0001

Baltes, P. B., and Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychol. Aging* 12, 12–21. doi: 10.1037//0882-7974.12.1.12

Barrouillet, P., and Camos, V. (2020). "The time-based resource-sharing model of working memory," in *Working memory*, eds P. Barrouillet and V. Camos (Oxford: Oxford University Press), doi: 10.1093/oso/9780198842286.003.0004

Bavelier, D., and Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nat. Rev. Neurosci.* 3, 443–452. doi: 10.1038/nrn848

Bernstein, L. E., Tucker, P. E., and Auer, E. T. (1998). Potential perceptual bases for successful use of a vibrotactile speech perception aid. *Scand. J. Psychol.* 39, 181–186. doi: 10.1111/1467-9450.393076

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—A review. *Trends Amplif.* 17, 75–93. doi: 10.1177/1084713813495459

Blomberg, R., Danielsson, H., Rudner, M., Söderlund, G. B. W., and Rönnberg, J. (2019). Speech processing difficulties in attention deficit hyperactivity disorder. *Front. Psychol.* 10:1536. doi: 10.3389/fpsyg.2019.01536

Bola, Ł, Zimmermann, M., Mostowski, P., Jednoróg, K., Marchewka, A., Rutkowski, P., et al. (2017). Task-specific reorganization of the auditory cortex in deaf humans. *Proc. Natl. Acad. Sci. U.S.A.* 114, E600–E609. doi: 10.1073/pnas.1609000114

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696

Bsharat-Maalouf, D., and Karawani, H. (2022). Learning and bilingualism in challenging listening conditions: How challenging can it be? *Cognition* 222:105018. doi: 10.1016/j.cognition.2022.105018

Calandruccio, L., Dhar, S., and Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *J. Acoust. Soc. Am.* 128, 860–869. doi: 10.1121/1.3458857

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science (New York, N.Y.)* 276, 593–596. doi: 10.1126/science.276.5312.593

Cardin, V., Grin, K., Vinogradova, V., and Manini, B. (2020). Crossmodal reorganisation in deafness: Mechanisms for functional preservation and functional change. *Neurosci. Biobehav. Rev.* 113, 227–237. doi: 10.1016/j.neubiorev.2020.03.019

Cardin, V., Orfanidou, E., Rönnberg, J., Capek, C. M., Rudner, M., and Woll, B. (2013). Dissociating cognitive and sensory-neural plasticity in human superior temporal cortex. *Nat. Commun.* 4:1473. doi: 10.1038/ncomms2463

Cardin, V., Rudner, M., De Oliveira, R. F., Andin, J., Su, M. T., Beese, L., et al. (2018). The organization of working memory networks is shaped by early sensory experience. *Cerebral Cortex.* 28, 3540–3554. doi: 10.1093/cercor/bhx222

Cowan, N. (2005). *Working memory capacity*. London: Psychology Press. doi: 10.4324/9780203342398

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

Daneman, M., and Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonom. Bull. Rev.* 3, 422–433. doi: 10.3758/BF03214546

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *J. Exp. Psychol.* 134, 222–241. doi: 10.1037/0096-3445.134.2.222

Dolležal, L.-V., Brechmann, A., Klump, G. M., and Deike, S. (2014). Evaluating auditory stream segregation of SAM tone sequences by subjective and objective psychoacoustical tasks, and brain activity. *Front. Neurosci.* 8:119. doi: 10.3389/fnins.2014.00119

Edwards, B. (2016). A model of auditory-cognitive processing and relevance to clinical applicability. *Ear Hear.* 37, 85S–91S. doi: 10.1097/AUD.0000000000000308

Ericsson, K. A., and Kintsch, W. (1995). Long-term working memory. *Psychol. Rev.* 102, 211–245. doi: 10.1037/0033-295X.102.2.211

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Fell, J., and Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nat. Rev. Neurosci.* 12, 105–118. doi: 10.1038/nrn2979

Fortunato, S., Forli, F., Guglielmi, V., De Corso, E., Paludetti, G., Berrettini, S., et al. (2016). A review of new insights on the association between hearing loss and cognitive decline in ageing. *Acta Otorhinolaryngol.* 36, 155–166. doi: 10.14639/0392-100X-993

Füllgrabe, C., Moore, B. C. J., and Stone, M. A. (2014). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6:347. doi: 10.3389/fnagi.2014.00347

Füllgrabe, C., and Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processingfor listeners with normal hearing thresholds. *Front. Psychol.* 7:1268. doi: 10.3389/fpsyg.2016.01268

Gatehouse, S., Naylor, G., and Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *Int. J. Audiol.* 42, S77–S85. doi: 10.3109/14992020309074627

Gatehouse, S., Naylor, G., and Elberling, C. (2006). Linear and nonlinear hearing aid fittings—1: Patterns of benefit. *Int. J. Audiol.* 45, 130–152. doi: 10.1080/14992020500429518

Gates, G. A. (2012). Central presbycusis: An emerging view. *Otolaryngol. Head Neck Surg.* 147, 1-2. doi: 10.1177/0194599812446282

Gates, G. A., Cobb, J. L., Linn, R. T., Rees, T., Wolf, P. A., and D'Agostino, R. B. (1996). Central auditory dysfunction, cognitive dysfunction, and dementia in older people. *Arch. Otolaryngol. Head Neck Surg.* 122, 161–167. doi: 10.1001/archotol.1996.01890140047010

Gibson, J. J. (1966). *The senses considered as perceptual systems*. London: GEORGE ALLEN & UNWIN LTD.

Giraud, A. L., Truy, E., and Frackowiak, R. (2001). Imaging plasticity in cochlear implant patients. *Audiol. Neuro Otol.* 6, 381–393. doi: 10.1159/000046847

Giraud, A. L., Truy, E., Frackowiak, R. S., Grégoire, M. C., Pujol, J. F., and Collet, L. (2000). Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients. *Brain* 123, 1391–1402. doi: 10.1093/brain/123.7.1391

Gordon-Salant, S., and Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear Hear.* 37, 593–602. doi: 10.1097/AUD.0000000000000316

Gray, R., Sarampalis, A., Başkent, D., and Harding, E. E. (2022). Working-memory, alpha-theta oscillations and musical training in older age: Research perspectives for speech-on-speech perception. *Front. Aging Neurosci.* 14:806439. doi: 10.3389/fnagi.2022.806439

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Percept. Psychophys.* 28, 267–283. doi: 10.3758/BF03204386

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scand. Audiol.* 11, 79-87. doi: 10.3109/01050398209076203

Hagoort, P., and Levinson, S. C. (2014). "Neuropragmatics," in *The cognitive neurosciences*, 5th Edn, eds M. S. Gazzaniga and G. R. Mangun (Cambridge, MA: MIT Press), 667–674.

Hällgren, M., Larsby, B., and Arlinger, S. (2006). A Swedish version of the hearing in noise test (HINT) for measurement of speech recognition. *Int. J. Audiol.* 45, 227–237. doi: 10.1080/14992020500429583

Han, M. K., Storkel, H., and Bontempo, D. E. (2019). The effect of neighborhood density on children's word learning in noise. *J. Child Lang.* 46, 153–169. doi: 10.1017/S0305000918000284

Hanna, J., Mejias, S., Schelstraete, M.-A., Pulvermüller, F., Shtyrov, Y., and Van der Lely, H. K. J. (2014). Early activation of Broca's area in grammar processing as revealed by the syntactic mismatch negativity and distributed source analysis. *Cogn. Neurosci.* 5, 66–76. doi: 10.1080/17588928.2013.860087

Hasting, A. S., Kotz, S. A., and Friederici, A. D. (2007). Setting the stage for automatic syntax processing: The mismatch negativity as an indicator of syntactic priming. *J. Cogn. Neurosci.* 19, 386–400. doi: 10.1162/jocn.2007.19.3.386

Heinrich, A., Henshaw, H., and Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Front. Psychol.* 6:782. doi: 10.3389/fpsyg.2015.00782

Heinrich, A., and Schneider, B. A. (2011). Elucidating the effects of ageing on remembering perceptually distorted word pairs. *Q. J. Exp. Psychol.* 64, 186–205. doi: 10.1080/17470218.2010.492621

Hewitt, D. (2017). Age-related hearing loss and cognitive decline: You haven't heard the half of it. *Front. Aging Neurosci.* 9:112. doi: 10.3389/fnagi.2017.00112

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev.* 8, 393–402. doi: 10.1038/nrn2113

Holmer, E., Heimann, M., and Rudner, M. (2016). Imitation, sign language skill and the developmental ease of language understanding (D-ELU) model. *Front Psychol.* 7:107. doi: 10.3389/fpsyg.2016.00107

Holmer, E., and Rudner, M. (2020). *Developmental ease of language understanding model and literacy acquisition: Evidence from deaf and hard-of-hearing signing children*. Washington, DC: Gallaudet University Press, 153–173.

Homman, L., Danielsson, H., and Rönnberg, J. (submitted). A structural equation mediation model captures the predictions amongst the parameters of the Ease of Language Understanding model. *Front. Psychol. Auditory Cogn. Neurosci.*

Hua, H., Johansson, B., Magnusson, L., Lyxell, B., and Ellis, R. J. (2017). Speech recognition and cognitive skills in bimodal cochlear implant users. *J. Speech Lang. Hear. Res.* 60, 2752–2763. doi: 10.1044/2017_JSLHR-H-16-0276

Humes, L. E. (2013). Understanding the speech-understanding problems of older adults. *Am. J. Audiol.* 22, 303–305. doi: 10.1044/1059-0889(2013/12-0066

Hygge, S., Rönnberg, J., Larsby, B., and Arlinger, S. (1992). Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. *J. Speech Hear. Res.* 35, 208–215. doi: 10.1044/jshr.3501.208

Janse, E., and Andringa, S. (2021). The roles of cognitive abilities and hearing acuity in older adults' recognition of words taken from fast and spectrally reduced speech. *Appl. Psycholinguist.* 42, 1–28. doi: 10.1017/S0142716421000047

Jones, G., Cabiddu, F., Andrews, M., and Rowland, C. (2021). Chunks of phonological knowledge play a significant role in children's word learning and explain effects of neighborhood size, phonotactic probability, word frequency and word length. *J. Memory Lang.* 119:104232. doi: 10.1016/j.jml.2021.104232

Jones, G., Gobet, F., and Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Dev. Sci.* 10, 853–873. doi: 10.1111/j.1467-7687.2007.00638.x

Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychol. Rev.* 99, 122–149. doi: 10.1037/0033-295x.99.1.122

Keidser, G., Seeto, M., Rudner, M., Hygge, S., and Rönnberg, J. (2015). On the relationship between functional hearing and depression. *Int. J. Audiol.* 54, 653–664. doi: 10.3109/14992027.2015.1046503

Kennedy-Higgins, D., Devlin, J. T., and Adank, P. (2020). Cognitive mechanisms underpinning successful perception of different speech distortions. *J. Acoust. Soc. Am.* 147, 2728–2740. doi: 10.1121/10.0001160

Kilman, L., Zekveld, A., Hällgren, M., and Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Front. Psychol.* 5:651. doi: 10.3389/fpsyg.2014.00651

Kral, A., and Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends Neurosci.* 35, 111–122. doi: 10.1016/j.tins.2011.09.004

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev.* 11, 599–605. doi: 10.1038/nrn2882

Kraus, N., Strait, D. L., and Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Ann. N Y Acad. Sci.* 1252, 100–107. doi: 10.1111/j.1749-6632.2012.06463.x

Kurthen, I., Meyer, M., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2020). Individual differences in peripheral hearing and cognition reveal sentence processing differences in healthy older adults. *Front. Neurosci.* 14:573513. doi: 10.3389/fnins.2020.573513

Levänen, S. (1998). Neuromagnetic studies of human auditory cortex function and reorganization. *Scand. Audiol. Suppl.* 49, 1–6. doi: 10.1080/010503998420595

Lidestam, B., Moradi, S., Pettersson, R., and Ricklefs, T. (2014). Audiovisual training is better than auditory-only training for auditory-only speech-in-noise identification. *J. Acoust. Soc. Am.* 136, EL142–EL147. doi: 10.1121/1.4890200

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734. doi: 10.1016/S0140-6736(17)31363-6

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001

Ludman, C. N., Summerfield, A. Q., Hall, D., Elliott, M., Foster, J., Hykin, J. L., et al. (2000). Lip-reading ability and patterns of cortical activation studied using fMRI. *Br. J. Audiol.* 34, 225–230. doi: 10.3109/03005364000000132

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *Int. J. Audiol.* 42, 49–58. doi: 10.3109/14992020309074624

Lunner, T., Rudner, M., and Rönnberg, J. (2009). Cognition and hearing aids. *Scand. J. Psychol.* 50, 395–403. doi: 10.1111/j.1467-9450.2009.00742.x

Lunner, T., and Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *J. Am. Acad. Audiol.* 18, 604–617. doi: 10.3766/jaaa.18.7.7

Lyxell, B. (1994). Skilled speechreading: A single-case study. *Scand. J. Psychol.* 35, 212–219. doi: 10.1111/j.1467-9450.1994.tb00945.x

Lyxell, B., Arlinger, S., Andersson, J., Harder, H., Näsström, E., Svensson, H., et al. (1996). Information-processing capabilities and cochlear implants. Pre-operative predictors for speech understanding. *J. Deaf Stud. Deaf Educ.* 11, 190–201. doi: 10.1093/oxfordjournals.deafed.a014294

Lyxell, B., and Rönnberg, J. (1987). Guessing and speechreading. *Br. J. Audiol.* 21, 13–20. doi: 10.3109/03005368709077769

MacSweeney, M., Campbell, R., Calvert, G. A., McGuire, P. K., David, A. S., Suckling, J., et al. (2001). Dispersed activation in the left temporal cortex for speech-reading in congenitally deaf people. *Proc. Biol. Sci.* 268, 451–457. doi: 10.1098/rspb.2000.0393

Marsh, J. E., and Campbell, T. A. (2016). Processing complex sounds passing through the rostral brainstem: The new early filter model. *Front. Neurosci.* 10:136. doi: 10.3389/fnins.2016.00136

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9

Marsja, E., Stenbäck, V., Moradi, S., Danielsson, H., and Rönnberg, J. (2022). Is having hearing loss fundamentally different? Multigroup structural equation modeling of the effect of cognitive functioning on speech identification. *Ear Hear.* doi: 10.1097/AUD.0000000000001196 [Epub ahead of print].

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Lang. Cogn. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 58, 22–33. doi: 10.1080/02724980443000151

Merabet, L. B., and Pascual-Leone, A. (2010). Neural reorganization following sensory loss: The opportunity of change. *Nat. Rev. Neurosci.* 11, 44–52. doi: 10.1038/nrn2758

Mishra, S., Shubhadarshan, A., Behera, D., and Sahoo, R. (2021). Can working memory capacity predict speech perception in presence of noise in older adults? *Int. J. Educ. Res.* 10, 18–23.

Moradi, S., Lidestam, B., Saremi, A., and Rönnberg, J. (2014b). Gated auditory speech perception: Effects of listening conditions and cognitive capacity. *Front. Psychol.* 5:531. doi: 10.3389/fpsyg.2014.00531

Moradi, S., Lidestam, B., Hällgren, M., and Rönnberg, J. (2014a). Gated auditory speech perception in elderly hearing aid users and elderly normal-hearing individuals: Effects of hearing impairment and cognitive capacity. *Trends Hear.* 18:2331216514545406. doi: 10.1177/2331216514545406

Moradi, S., Lidestam, B., Ng, E. H. N., Danielsson, H., and Rönnberg, J. (2019). Perceptual doping: An audiovisual facilitation effect on auditory speech processing, from phonetic feature extraction to sentence identification in noise. *Ear Hear.* 40, 312–327. doi: 10.1097/AUD.0000000000000616

Moradi, S., Lidestam, B., and Rönnberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. *Front. Psychol.* 4:359. doi: 10.3389/fpsyg.2013.00359

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear Hear.* 16, 6–18.

Näätänen, R., and Escera, C. (2000). Mismatch negativity: Clinical and other applications. *Audiol. Neuro Otol.* 5, 105–110. doi: 10.1159/000013874

Ng, E. H. N., and Rönnberg, J. (2020). Hearing aid experience and background noise affect the robust relationship between working memory and speech recognition in noise. *Int. J. Audiol.* 59, 208–218. doi: 10.1080/14992027.2019.1677951

Ng, E. H., Rudner, M., Lunner, T., Pedersen, M. S., and Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int. J. Audiol.* 52, 433–441. doi: 10.3109/14992027.2013.776181

Ng, E. H. N., Rudner, M., Lunner, T., and Rönnberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear Hear.* 36, 82–91. doi: 10.1097/AUD.0000000000000080

Nilsson, L.-G., Bäckman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., et al. (1997). The Betula prospective cohort study: Memory, health, and aging. *Aging Neuropsychol. Cogn.* 4, 1–32. doi: 10.1080/13825589708256633

Obleser, J., Wostmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* 32, 12376–12383. doi: 10.1523/JNEUROSCI.4908-11.2012

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* 39, 204–214. doi: 10.1097/AUD.0000000000000494

Petersen, E. B., Wöstmann, M., Obleser, J., Stenfelt, S., and Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Front. Psychol.* 6:177. doi: 10.3389/fpsyg.2015.00177

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *Int. J. Audiol.* 42, 59–67. doi: 10.3109/14992020309074625

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Powell, D. S., Oh, E. S., Reed, N. S., Lin, F. R., and Deal, J. A. (2021). Hearing loss and cognition: What we know and where we need to go. *Front. Aging Neurosci.* 13:769405. doi: 10.3389/fnagi.2021.769405

Richter, M. (2013). A closer look into the multi-layer structure of motivational intensity theory. *Soc. Pers. Psychol. Compass* 7, 1–12. doi: 10.1111/spc3.12007

Rönnberg, J. (1990). Cognitive and communicative function: The effects of chronological age and "handicap age". *Eur. J. Cogn. Psychol.* 2, 253–273. doi: 10.1080/09541449008406207

Rönnberg, J. (1993). Cognitive characteristics of skilled tactiling: The case of GS. *Eur. J. Cogn. Psychol.* 5, 19–33. doi: 10.1080/09541449308406512

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *Int. J. Audiol.* 42, S68–S76. doi: 10.3109/14992020309074626

Rönnberg, J., Andersson, J., Andersson, U., Johansson, K., Lyxell, B., and Samuelsson, S. (1998). Cognition as a bridge between signal and dialogue: Communication in the hearing impaired and deaf. *Scand. Audiol.* 27, 101–108. doi: 10.1080/010503998420720

Rönnberg, J., Andersson, J., Samuelsson, S., Söderfeldt, B., Lyxell, B., and Risberg, J. (1999). A speechreading expert: The case of MM. *J. Speech Lang. Hear. Res.* 42, 5–20. doi: 10.1044/jslhr.4201.05

Rönnberg, J., Danielsson, H., Rudner, M., Arlinger, S., Sternäng, O., Wahlin, A., et al. (2011). Hearing loss is negatively related to episodic and semantic long-term memory but not to short-term memory. *J. Speech Lang. Hear. Res.* 54, 705–726. doi: 10.1044/1092-4388(2010/09-0088

Rönnberg, J., Holmer, E., and Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Int. J. Audiol.* 58, 247–261. doi: 10.1080/14992027.2018.1551631

Rönnberg, J., Holmer, E., and Rudner, M. (2021). Cognitive hearing science: Three memory systems, two approaches, and the ease of language understanding model. *J. Speech Lang. Hear. Res.* 64, 359–370. doi: 10.1044/2020_JSLHR-20-00007

Rönnberg, J., Holmer, E., and Rudner, M. (2022). "Working memory and the ease of language understanding model," in *The Cambridge handbook of working memory*, eds W. Schwieter and Z. E. Wen (Cambridge: Cambridge University Press), 197–218. doi: 10.1017/9781108955638.013

Rönnberg, J., Hygge, S., Keidser, G., and Rudner, M. (2014). The effect of functional hearing loss and age on long- and short-term visuospatial memory: Evidence from the UK biobank resource. *Front. Aging Neurosci.* 6:326. doi: 10.3389/fnagi.2014.00326

Rönnberg, J., Lunner, T., Ng, E. H. N., Lidestam, B., Zekveld, A. A., Sörqvist, P., et al. (2016). Hearing impairment, cognition and speech understanding: Exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *Int. J. Audiol.* 55, 623–642. doi: 10.1080/14992027.2016.1219775

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Rönnberg, J., Öhngren, G., and Nilsson, L. G. (1982). Hearing deficiency, speechreading and memory functions. *Scand. Audiol.* 11, 261–268. doi: 10.3109/01050398209087477

Rönnberg, J., Öhngren, G., and Nilsson, L. G. (1983). Speechreading performance evaluated by means of TV and real-life presentation. A comparison between a normally hearing, moderately and profoundly hearing-impaired group. *Scand. Audiol.* 12, 71–77. doi: 10.3109/01050398309076227

Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 47, S99–S105. doi: 10.1080/14992020802301167

Rönnberg, J., Rudner, M., and Ingvar, M. (2004). Neural correlates of working memory for sign language. *Brain Res. Cogn. Brain Res.* 20, 165–182. doi: 10.1016/j.cogbrainres.2004.03.002

Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise Health* 12:263. doi: 10.4103/1463-1741.70505

Rudner, M., Foo, C., Rönnberg, J., and Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scand. J. Psychol.* 50, 405–418. doi: 10.1111/j.1467-9450.2009.00745.x

Rudner, M., Foo, C., Sundewall-Thorén, E., Lunner, T., and Rönnberg, J. (2008). Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. *Int. J. Audiol.* 47, S91–S98. doi: 10.1080/14992020802304393

Rudner, M., Fransson, P., Ingvar, M., Nyberg, L., and Rönnberg, J. (2007). Neural representation of binding lexical signs and words in the episodic buffer of working memory. *Neuropsychologia* 45, 2258–2276. doi: 10.1016/j.neuropsychologia.2007.02.017

Rudner, M., Seeto, M., Keidser, G., Johnson, B., and Rönnberg, J. (2019). Poorer speech reception threshold in noise is associated with lower brain volume in auditory and cognitive processing regions. *J. Speech Lang. Hear. Res.* 62, 1117–1130. doi: 10.1044/2018_JSLHR-H-ASCC7-18-0142

Samuelsson, S., and Rönnberg, J. (1993). Implicit and explicit use of scripted constraints in lip-reading. *Eur. J. Cogn. Psychol.* 5, 201–233. doi: 10.1080/09541449308520116

Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J. Speech Lang. Hear. Res.* 52, 1230–1240. doi: 10.1044/1092-4388(2009/08-0111

Signoret, C., Andersen, L. M., Dahlström, Ö, Blomberg, R., Lundqvist, D., Rudner, M., et al. (2020). The influence of form- and meaning-based predictions on cortical speech processing under challenging listening conditions: A MEG study. *Front. Neurosci.* 14:573254. doi: 10.3389/fnins.2020.573254

Signoret, C., Gaudrain, E., and Perrin, F. (2013). Similarities in the neural signature for the processing of behaviorally categorized and uncategorized speech sounds. *Eur. J. Neurosci.* 37, 777–785. doi: 10.1111/ejn.12097

Signoret, C., Johnsrude, I., Classon, E., and Rudner, M. (2018). Combined effects of form- and meaning-based predictability on perceived clarity of speech. *Hum. Percept. Perform.* 44, 277–285. doi: 10.1037/xhp0000442

Signoret, C., and Rudner, M. (2019). Hearing impairment and perceived clarity of predictable speech. *Ear Hear.* 40, 1140–1148. doi: 10.1097/AUD.0000000000000689

Silvestrini, N., Musslick, S., Berry, A. S., and Vassena, E. (2022). An integrative effort: Bridging motivational intensity theory and recent neurocomputational and neuronal models of effort and control allocation. *Psychol. Rev.* doi: 10.1037/rev0000372

Söderfeldt, B., Ingvar, M., Rönnberg, J., Eriksson, L., Serrander, M., and Stone-Elander, S. (1997). Signed and spoken language perception studied by positron emission tomography. *Neurology* 49, 82–87. doi: 10.1212/wnl.49.1.82

Söderfeldt, B., Rönnberg, J., and Risberg, J. (1994). Regional cerebral blood flow in sign language users. *Brain* 46, 59-68. doi: 10.1006/brln.1994.1004

Sörqvist, P., and Rönnberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: What is the role for working memory capacity? *J. Speech Lang. Hear. Res.* 55, 210–218. doi: 10.1044/1092-4388(2011/10-0353

Sörqvist, P., Stenfelt, S., and Rönnberg, J. (2012). Working memory capacity and visual-verbal cognitive load modulate auditory-sensory gating in the brainstem: Toward a unified view of attention. *J. Cogn. Neurosci.* 24, 2147–2154. doi: 10.1162/jocn_a_00275

Souza, P., and Arehart, K. (2015). Robust relationship between reading span and speech recognition in noise. *Int. J. Audiol.* 54, 705–713. doi: 10.3109/14992027.2015.1043062

Souza, P., Arehart, K., Schoof, T., Anderson, M., Strori, D., and Balmert, L. (2019). Understanding variability in individual response to hearing aid signal processing in wearable hearing aids. *Ear Hear.* 40, 1280–1292. doi: 10.1097/AUD.0000000000000717

Souza, P., and Sirow, L. (2014). Relating working memory to compression parameters in clinically fit hearing AIDS. *Am. J. Audiol.* 23, 394–401. doi: 10.1044/2014_AJA-14-0006

Stenbäck, V., Hällgren, M., Lyxell, B., and Larsby, B. (2015). The Swedish hayling task, and its relation to working memory, verbal ability, and speech-recognition-in-noise. *Scand. J. Psychol.* 56, 264–272. doi: 10.1111/sjop.12206

Stenbäck, V., Marsja, E., Hällgren, M., Lyxell, B., and Larsby, B. (2021). The contribution of age, working memory capacity, and inhibitory control on speech recognition in noise in young and older adult listeners (world). *J Speech Lang Hear Res.* 64, 4513–4523. doi: 10.1044/2021_JSLHR-20-00251

Stenfelt, S., and Rönnberg, J. (2009). The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scand. J. Psychol.* 50, 385–393. doi: 10.1111/j.1467-9450.2009.00748.X

Tillberg, I., Rönnberg, J., Svärd, I., and Ahlner, B. (1996). Audio-visual speechreading in a group of hearing aid users. The effects of onset age, handicap age, and degree of hearing loss. *Scand. Audiol.* 25, 267–272. doi: 10.3109/01050399609074966

Tse, C.-Y., Rinne, T., Ng, K. K., and Penney, T. B. (2013). The functional role of the frontal cortex in pre-attentive auditory change detection. *Neuroimage* 83, 870–879. doi: 10.1016/j.neuroimage.2013.07.037

Tun, P. A., McCoy, S., and Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychol. Aging* 24, 761–766. doi: 10.1037/a0014802

Twomey, T., Waters, D., Price, C. J., Evans, S., and MacSweeney, M. (2017). How auditory experience differentially influences the function of left and right superior temporal cortices. *J. Neurosci.* 37, 9564–9573. doi: 10.1523/JNEUROSCI.0846-17.2017

Uchida, Y., Nishita, Y., Otsuka, R., Sugiura, S., Sone, M., Yamasoba, T., et al. (2021). Aging brain and hearing: A mini-review. *Front. Aging Neurosci.* 13:791604. doi: 10.3389/fnagi.2021.791604

Usrey, W. M., and Sherman, S. M. (2019). Corticofugal circuits: Communication lines from the cortex to the rest of the brain. *J. Comp. Neurol.* 527, 640–650. doi: 10.1002/cne.24423

Van Engen, K. J., and Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *J. Acoust. Soc. Am.* 121, 519–526. doi: 10.1121/1.2400666

Vermeire, K., Knoop, A., De Sloovere, M., Bosch, P., and van den Noort, M. (2019). Relationship between working memory and speech-in-noise recognition in young and older adult listeners with age-appropriate hearing. *J. Speech Lang. Hear. Res.* 62, 3545–3553. doi: 10.1044/2019_JSLHR-H-18-0307

Wayne, R. V., and Johnsrude, I. S. (2015). A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Res. Rev.* 23, 154–166. doi: 10.1016/j.arr.2015.06.002

Younan, D., Petkus, A. J., Widaman, K. F., Wang, X., Casanova, R., Espeland, M. A., et al. (2020). Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain* 143, 289–302. doi: 10.1093/brain/awz348

Zatorre, R. J. (2001). Do you see what I'm saying? Interactions between auditory and visual cortices in cochlear implant users. *Neuron* 31, 13–14. doi: 10.1016/s0896-6273(01)00347-6

Zekveld, A. A., Pronk, M., Danielsson, H., and Rönnberg, J. (2018). Reading behind the lines: The factors affecting the text reception threshold in hearing aid users. *J. Speech Lang. Hear. Res.* 61, 762–775. doi: 10.1044/2017_JSLHR-H-17-0196

Zekveld, A. A., Rudner, M., Johnsrude, I. S., and Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *J. Acoust. Soc. Am.* 134, 2225–2234. doi: 10.1121/1.4817926