The Contribution of Auditory and Cognitive Factors to Intelligibility of Words and Sentences in Noise

Antje Heinrich and Sarah Knight

Abstract Understanding the causes for speech-in-noise (SiN) perception difficulties is complex, and is made even more difficult by the fact that listening situations can vary widely in target and background sounds. While there is general agreement that both auditory and cognitive factors are important, their exact relationship to SiN perception across various listening situations remains unclear. This study manipulated the characteristics of the listening situation in two ways: first, target stimuli were either isolated words, or words heard in the context of low- (LP) and high-predictability (HP) sentences; second, the background sound, speech-modulated noise, was presented at two signal-to-noise ratios. Speech intelligibility was measured for 30 older listeners (aged 62–84) with age-normal hearing and related to individual differences in cognition (working memory, inhibition and linguistic skills) and hearing ($PTA_{0.25–8\,kHz}$ and temporal processing). The results showed that while the effect of hearing thresholds on intelligibility was rather uniform, the influence of cognitive abilities was more specific to a certain listening situation. By revealing a complex picture of relationships between intelligibility and cognition, these results may help us understand some of the inconsistencies in the literature as regards cognitive contributions to speech perception.

Keywords Speech-in-noise perception · Listening situations · Cognitive and auditory variables

1 Introduction

Speech-in-noise (SiN) perception is something that many listener groups, including older adults, find difficult. Previous research has shown that hearing sensitivity cannot account for all speech perception difficulties, particularly in noise (Schneider

A. Heinrich (✉) · S. Knight
MRC Institute of Hearing Research, Nottingham NG7 2RD, UK
e-mail: antje.heinrich@ihr.mrc.ac.uk

S. Knight
e-mail: sarah.knight@ihr.mrc.ac.uk

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing,* Advances in Experimental Medicine and Biology 894,
DOI 10.1007/978-3-319-25474-6_5

and Pichora-Fuller 2000; Wingfield and Tun 2007). Consequently, cognition has emerged as another key factor. While there is general agreement that a relationship between cognition and speech perception exists, its nature and extent remain unclear. No single cognitive component has emerged as being important for all listening contexts, although working memory, as tested by reading span, often appears to be important.

Working memory (WM) has no universally-accepted definition. One characterisation posits that WM refers to the ability to simultaneously store and process task-relevant information (Daneman and Carpenter 1980). WM tasks may emphasize either storage or processing. Storage-heavy tasks, such as Digit Span and Letter-Number Sequencing (Wechsler 1997), require participants to repeat back material either unchanged or slightly changed. Processing-heavy tasks, such as the Reading Span task (Daneman and Carpenter 1980), require a response that differs considerably from the original material and is only achieved by substantial mental manipulation.

The correlation between WM and speech perception, particularly in noise, tends to be larger when the WM task is complex i.e. processing-heavy (Akeroyd 2008). However, this is only a general trend: not all studies show the expected correlation (Koelewijn et al. 2012), and some studies show significant correlations between WM and SiN perception even though the WM measure was storage-, not processing-heavy (Humes et al. 2006; Rudner et al. 2008). Why these inconsistencies occurred remains to be understood.

WM can also be conceptualised in terms of inhibition of irrelevant information (Engle and Kane 2003), which has again been linked to speech perception. For instance, poor inhibition appears to increase susceptibility to background noise during SiN tasks (Janse 2012). Finally, general linguistic competence—specifically vocabulary size and reading comprehension—has also been shown to aid SiN perception in some situations (Avivi-Reich et al. 2014).

Some of the inconsistencies in the relationship between SiN perception and cognition are most likely caused by varying combinations of speech perception and cognitive tasks. Like cognitive tasks, speech perception tasks can be conceptualised in different ways. Speech tasks may vary along several dimensions including the complexity of the target (e.g. phonemes vs. sentences), of the background signal (e.g. silence vs. steady-state noise vs. babble), and/or the overall difficulty (e.g. low vs. high signal-to-noise ratio) of the listening situation. It is difficult to know if and to what extent variations along these dimensions affect the relationship to cognition, and whether these variations may explain, at least in part, inconsistencies between studies. For instance, could differences in target speech help explain why some studies (Desjardins and Doherty 2013; Moradi et al. 2014), using a complex sentence perception test, found a significant correlation between reading span and intelligibility, while another study (Kempe et al. 2012), using syllables and the same cognitive test, did not? The goal of this study is to systematically vary the complexity of the listening situation and to investigate how this variation affects the relationship between intelligibility and assorted cognitive measures.

Finally, it is important to note that a focus on cognitive contributions does not imply that auditory contributions to SiN perception are unimportant. Besides over-

all hearing sensitivity we also obtained a suprathreshold measure of temporal processing by measuring the sensitivity to change in interaural correlation. This task is assumed to estimate loss of neural synchrony in the auditory system (Wang et al. 2011), which in turn has been suggested to affect SiN perception.

2 Methods

2.1 Listeners

Listeners were 30 adults aged over 60 (mean: 70.2 years, SD: 6.7, range: 62–84) with age-normal hearing. Exclusion criteria were hearing aid use and non-native English language status.

2.2 Tasks

2.2.1 Speech Tasks

Sentences Stimuli were 112 sentences from a recently developed sentence pairs test (Heinrich et al. 2014). This test, based on the SPIN-R test (Bilger et al. 1984), comprises sentence pairs with identical sentence-final monosyllabic words, which are more or less predictable from the preceding context (e.g. 'We'll never get there at this rate' versus 'He's always had it at this rate'). High and low predictability (HP/LP) sentence pairs were matched for duration, stress pattern, and semantic complexity, and were spoken by a male Standard British English speaker.

Words A total of 200 words comprising the 112 final words from the sentence task and an additional 88 monosyllables were re-recorded using a different male Standard British English speaker.

Noise All speech stimuli were presented in speech-modulated noise (SMN) derived from the input spectrum of the sentences themselves. Words were presented at signal-to-noise ratios (SNRs) of + 1 and − 2 dB, sentences at − 4 and − 7 dB. SNR levels were chosen to vary the overall difficulty of the task between 20 and 80 % accuracy. Intelligibility for each of six conditions (words and LP/HP sentences at low and high SNRs) was measured.

2.2.2 Auditory Task

Temporal Processing Task (TPT) Duration thresholds were obtained for detecting a change in interaural correlation (from 0 to 1) in the initial portion of a 1-s broad-

band (0–10 kHz) noise presented simultaneously to both ears. A three-down, one-up 2AFC procedure with 12 reversals, of which the last eight were used to estimate the threshold, was used. Estimates were based on the geometric mean of all collected thresholds (minimum 3, maximum 5).

2.2.3 Cognitive Tasks

Letter-Number Sequencing (LNS) The LNS (Wechsler 1997) measures mainly the storage component of WM although some manipulation is required. Participants heard a combination of numbers and letters and were asked to recall the numbers in ascending order, then the letters in alphabetical order. The number of items per trial increased by one every three trials; the task stopped when all three trials of a given length were repeated incorrectly. The outcome measure was the number of correct trials.

Reading Span Task (RST) The RST places greater emphasis on manipulation (Daneman and Carpenter 1980). In each trial participants read aloud unconnected complex sentences of variable length and recalled the final word of each sentence at the end of a trial. The number of sentences per trial increased by one every five trials, from two to five sentences. All participants started with trials of two sentences and completed all 25 trials. The outcome measure was the overall number of correctly recalled words.

Visual Stroop In a variant on the original Stroop colour/word interference task (Stroop 1935) participants were presented with grids of six rows of eight coloured blocks. In two grids (control), each of the 48 blocks contained "XXXX" printed in 20pt font at the centre of the block; in another two grids (experimental), the blocks contained a mismatched colour word (e.g. "RED" on a green background). In both cases, participants were asked to name the colour of each background block as quickly and accurately as possible. Interference was calculated by subtracting the time taken to name the colours on the control grids from the time taken to name the colours in the mismatched experimental grids.

Mill Hill Vocabulary Scale (MH) The Mill Hill (Raven et al. 1982) measures acquired verbal knowledge in a 20-word multiple-choice test. For each word participants selected the correct synonym from a list of six alternatives. A summary score of all correct answers was used.

Nelson-Denny Reading Test (ND) The Nelson-Denny (Brown et al. 1981) is a reading comprehension test containing eight short passages. Participants were given 20 min to read the passages and answer 36 multiple-choice questions. The outcome measure was the number of correctly answered questions.

2.3 Procedure

Testing was carried out in a sound-attenuated chamber using Sennheiser HD280 headphones. With the exception of the TPT all testing was in the left ear only. Testing took place over the course of two sessions around a week apart. Average pure-tone air-conduction thresholds (left ear) are shown in Fig. 1a. The pure-tone average (PTA) of all measured frequencies was used as an individual measure of hearing sensitivity. In order to determine the presentation level for all auditory stimuli including cognitive tasks, speech reception thresholds were obtained using 30 sentences from the Adaptive Sentence List (MacLeod and Summerfield 1990). The speech level was adaptively varied starting at 60 dB SPL and the average presentation level of the last two reversals of a three-down, one-up paradigm with a 2 dB step size was used. Presenting all stimuli at 30 dB above that level was expected to account for differences in intelligibility in quiet.

TPT and cognitive tasks were split between sessions without a strict order. The word and sentence tasks were always tested in different sessions, with the order of word and sentence tasks counterbalanced across participants. Each sentence-final word was only heard once, either in the context of an HP or an LP sentence, and half the sentences of each type were heard with high or low SNR. Across all listeners, each sentence-final word was heard an equal number of times in all four conditions (sentence type × SNR). After hearing each sentence or word participants repeated as much as they could. Testing was self-paced. The testing set-up for the cognitive tasks was similar but adapted to task requirements.

3 Results

Figure 1b presents mean intelligibility for stimulus type and SNR A 3 stimulus type (words, LP sentences, HP sentences) by two SNR (high, low) repeated-measures ANOVA showed main effects of type ($F(2, 50) = 192.55$ $p < 0.001$, LP < words < HP) and SNR ($F(1, 25) = 103.43$, $p < 0.001$) but no interaction ($F(2, 50) = 1.78$, $p = 0.18$), and suggested that a 3-dB decrease in SNR reduced intelligibility for all three stimuli types by a similar amount (12%). There was also a significant difference in intelligibility between LP and HP sentences.

The effect of auditory and cognitive factors on intelligibility was examined in a series of separate linear mixed models (LMM with SSTYPE1). Each model included one auditory or cognitive variable, both as main effect and in its interactions with stimulus type (words, LP, HP) and SNR (L, H). The previously confirmed main effects for stimulus type and SNR were modelled but are not separately reported. Participants were included with random intercepts. Table 1 A displays the *p*-values for all significant effects. Non-significant results are not reported. Table 1 B displays bivariate correlations between each variable and the scores in each listening situation to aid interpretation. Note however that bivariate correlations do not need to be significant by themselves to drive a significant effect in an LMM.

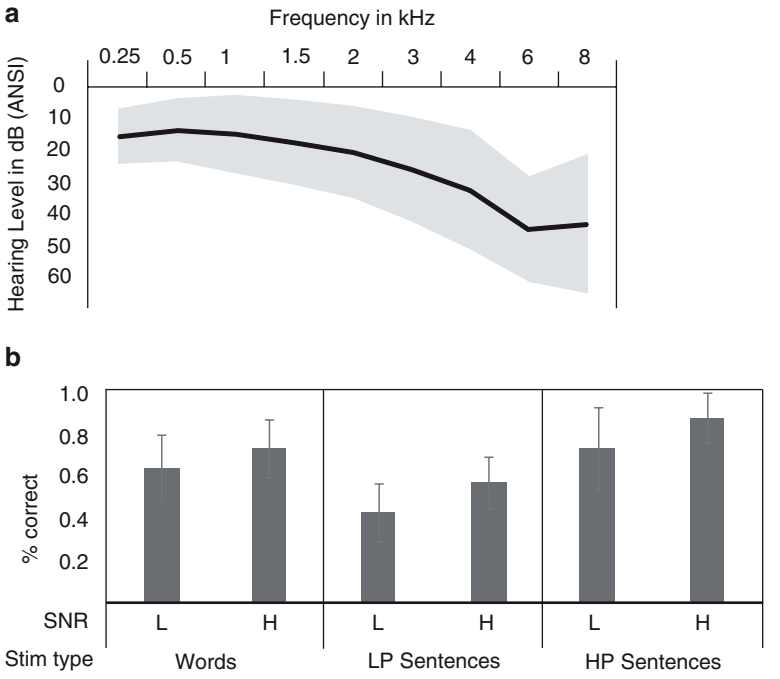

Fig. 1 **a** Audiometric thresholds (mean ± 1 SD) as a function of frequency for the left ear. **b** Intelligibility (mean + 1SD) for three stimuli types (words, LP/HP sentences) presented at high (*H*) and low (*L*) SNR

Table 1 A: Significant *p*-values in linear mixed models estimating the effect of each auditory and cognitive variable on listening situations varying in type of target speech (words, LP/HP sentences) and SNR. *PTA* pure-tone average 0.25–8 kHz (left ear), *LNS* letter-number sequencing, *RST* reading span task, *ND* Nelson-Denny reading comprehension, *ME* main effect. 1B: Pearson product-moment correlations between each auditory/cognitive variable and intelligibility in each of six listening situations. *Words* monosyllables, *LP* low-predictability, *HP* high-predictability. Significant correlations are in italics

		A. p-values			B. Pearson product-moment correlations					
					Words		LP sentences		HP sentences	
	var	ME	SNR *var	Type *SNR *var	Low SNR	High SNR	Low SNR	High SNR	Low SNR	High SNR
Auditory	PTA	<0.001	0.007	0.05	*−0.57*	*−0.60*	*−0.66*	*−0.53*	*−0.69*	*−0.61*
	TPT			0.04	−0.18	*−0.38*	−0.33	−0.20	−0.29	−0.26
Storage	LNS	0.003			0.34	*0.43*	*0.54*	*0.55*	0.38	*0.55*
Processing	RST		0.05		0.16	0.20	0.33	0.03	0.36	0.27
Inhibition	Stroop				0.01	0.06	−0.08	0.16	0.10	0.12
Vocab	MH				−0.03	−0.06	0.07	0.03	−0.02	0.22
Compr	ND				0.06	0.10	0.30	0.02	0.17	0.26

The main effect of PTA reflected the fact that average audiometric threshold was negatively correlated with intelligibility in all listening situations. The interaction with SNR occurred because the correlations tended to be greater for low SNRs than high SNRs. This was particularly true for the sentence stimuli leading to the three-way interaction (Type * SNR * PTA). The Type*SNR*TPT interaction occurred because TPT thresholds showed a significant negative correlation with intelligibility for word stimuli at high SNRs. The main effect of LNS occurred because a better storage-based WM score was beneficial for intelligibility in all tested listening situations. Examining the bivariate correlations to understand the interaction between WM processing-heavy RST and SNR suggests that a good score in the RST was most beneficial for listening situations with low SNR, although this relationship did not reach significance in any case. Intelligibility was not affected by inhibition, vocabulary size or reading comprehension abilities.

4 Discussion

In this study we assessed individual differences in auditory and cognitive abilities and investigated their predictive value for SiN perception across a range of tasks. By systematically varying the characteristics of the listening situation we hoped to resolve inconsistencies in the cognitive speech literature regarding correlations between cognitive and speech tasks. By assessing multiple relevant abilities we also aimed to understand if and how the contribution of a particular ability varied across listening situation.

The results suggest that the importance of a particular variable often, but not always, depends on the listening situation. Hearing sensitivity (PTA) and a basic WM task correlated with intelligibility in all tested situations. The results for PTA are somewhat surprising given that all speech testing was done at sensitivity-adjusted levels, which might have been expected to equate for PTA differences. The PTA measure may therefore capture some aspect of hearing that is not well represented by SRT.

The importance of WM for intelligibility is less surprising. However, the rather basic, storage-based WM task appeared to capture a more general benefit, at least in the tested listening situations, than the often-used RST. While the RST did predict intelligibility, the effect was stronger for acoustically difficult listening situations (low SNR). The data provide support for the notion that the relationship between RST and speech perception depends on the listening situations and that inconsistent results in the literature may have occurred because not all speech tasks engage WM processes enough to lead to a reliable correlation. Lastly, we showed an effect of individual differences in temporal processing on intelligibility, but this effect was limited to easily perceptible (high SNR) single words. Possibly, temporal information is most useful when the stimulus is clearly audible and no other semantic information is available.

These results suggest that different variables modulate listening in different situations, and that listeners may vary not only in their overall level of performance but also in how well they perceive speech in a particular situation depending on which auditory/cognitive abilities underpin listening in that situation and how successful the listener is at employing them.

Acknowledgments This research was supported by BBSRC grant BB/K021508/1.

References

Akeroyd MA (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(2):53–71

Avivi-Reich M, Daneman M, Schneider BA (2014) How age and linguistic competence alter the interplay of perceptual and cognitive factors when listening to conversations in a noisy environment. Front Syst Neurosci. doi:10.3389/fnsys.2014.00021

Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C (1984) Standardization of a test of speech perception in noise. J Speech Hear Res 27:32–48

Brown JI, Bennett JM, Hanna G (1981) The Nelson-Denny reading test. Riverside, Chicago

Daneman M, Carpenter PA (1980) Individual differences in working memory and reading. J Verbal Learning Verbal Behav 19:450–466

Desjardins JL, Doherty KA (2013) Age-related changes in listening effort for various types of masker noises. Ear Hear 34(3):261–272

Engle RW, Kane MJ (2003) Executive attention, working memory capacity, and a two-factor theory of cognitive control. In: Ross B (ed) Psychology of learning and motivation, vol 44. Elsevier, San Diego, pp 145–199

Heinrich A, Knight S, Young M, Moorhouse R, Barry J (2014) Assessing the effects of semantic context and background noise for speech perception with a new British English sentences set test (BESST). Paper presented at the British Society of Audiology Annual Conference, Keele University, UK

Humes LE, Lee JH, Coughlin MP (2006) Auditory measures of selective and divided attention in young and older adults using single-talker competition. J Acoust Soc Am 120(5):2926–2937. doi:10.1121/1.2354070

Janse E (2012) A non-auditory measure of interference predicts distraction by competing speech in older adults. Aging Neuropsychol Cogn 19:741–758. doi:10.1080/13825585.2011.652590

Kempe V, Thoresen JC, Kirk NW, Schaeffler F, Brooks PJ (2012) Individual differences in the discrimination of novel speech sounds: effects of sex, temporal processing, musical and cognitive abilities. PloS ONE 7(11):e48623. doi:10.1371/journal.pone.0048623

Koelewijn T, Zekveld AA, Festen JM, Ronnberg J, Kramer SE (2012) Processing load induced by informational masking is related to linguistic abilities. Int J Otolaryngol 1–11. doi:10.1155/2012/865731

MacLeod A, Summerfield Q (1990) A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. Br J Audiol 24:29–43

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

Raven JC, Raven J, Court JH (1982) Mill Hill vocabulary scale. Oxford Psychologists Press, Oxford

Rudner M, Foo C, Sundewall-Thorén E, Lunner T, Rönnberg J (2008) Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. Int J Audiol 47(2):91–98. doi:10.1080/14992020802304393

Schneider BA, Pichora-Fuller MK (2000) Implications of perceptual deterioration for cognitive aging research. In: Craik FIM, Salthouse TA (eds) The handbook of aging and cognition. Erlbaum, Mahwah, pp 155–219

Stroop JR (1935) Studies of interference in serial verbal reactions. J Exp Psychol 18:643–662

Wang M, Wu X, Li L, Schneider BA (2011) The effects of age and interaural delay on detecting a change in interaural correlation: the role of temporal jitter. Hear Res 275:139–149

Wechsler D (1997) Wechsler adult intelligence scale-3rd edition (WAIS-3®). Harcourt Assessment, San Antonio

Wingfield A, Tun PA (2007) Cognitive supports and cognitive constraints on comprehension of spoken language. J Am Acad Audiol 18:548–558