Cognitive factors in speech perception

A.A. Zekveld, S.E. Kramer, T. Houtgast
VU University medical center, The Netherlands, Email: aa.zekveld@vumc.nl

Abstract

Both auditory, modality-specific and cognitive, modality-aspecific functions influence the ability to perceive and comprehend speech in noisy listening situations. Important cognitive factors are working memory capacity and the ability to use linguistic context to fill in incomprehensible 'gaps' in the speech. When listening conditions are challenging (due to background noise or hearing loss), listeners rely to a greater extent on effortful top-down processes to improve speech comprehension. Therefore, it is important to measure both auditory and cognitive abilities when examining the causes of speech comprehension problems.

To prevent the performance on cognitive tests from being confounded by hearing loss, *visual tests* of relevant cognitive functions need to be considered when testing hearing impaired participants. Therefore, our group recently has developed the Text Reception Threshold (TRT) test, a visual analogue of the Speech Reception Threshold test. Several studies showed that speech comprehension in noise is associated with the TRT, indicating that both abilities rely on similar modality-aspecific cognitive processes. Future studies should focus on further unraveling the auditory and cognitive functions in speech comprehension and on the development of a clinically applicable test of relevant cognitive functions in speech comprehension in noise.

Introduction

Research over the last decades has shown that speech comprehension in difficult listening conditions depends on both peripheral *bottom-up* factors and modality-aspecific *top-down* cognitive abilities. The present paper gives a short overview of the cognitive abilities relevant for speech comprehension and the influence of age on these abilities. This paper furthermore describes a test measuring one of these abilities, verbal inference making, and it explains the value of applying physiological measures in studies to cognitive processes in speech comprehension.

Cognitive functions relevant for speech comprehension

Although sensory hearing loss is the main cause of speech communication difficulties [1, 2], the standard pure-tone audiogram and psycho-acoustical tests do not fully account for interindividual differences in the ability to comprehend speech in noise [2, 3, 4, 5]. Individuals with similar types and degrees of peripheral hearing loss can vary considerably in their ability to understand speech in noise. Part of this variance in speech comprehension can be explained by non-auditory functions like the ability to use context information to complete a partially comprehended sentence (e.g., [6, 7]; for a recent overview of relevant studies, see [8]). In 'analysis-by-synthesis' models of speech perception, correct comprehension is based the interaction between the acoustical analysis of the speech signal and cognitive processes that use the contextual information to synthesize a likely candidate for the phonetic speech segments [9].

Working memory [10, 11] is one of the cognitive abilities important for language comprehension. It is relevant for the processing and maintenance of incoming information in a limited-capacity 'store' in which the information can be manipulated. Specifically, the performance on *verbal* working memory tests (e.g., the reading span test, [12]) is related to speech comprehension (e.g., [8, 13]). Age-related declines in speech processing increase the deployment of working memory capacity during speech comprehension [14, 15, 16].

The ability to make use of linguistic context is also relevant for speech comprehension [1, 14, 17, 18, 19, 20, 21]. Individuals who are more skilled at using the linguistic context to fill in 'missing' auditory information will likely experience fewer problems in speech communication.

Top-down processes and ageing

Increasing age is associated with decreasing working memory functioning and lower processing speed [22, 23]. Other abilities, like word vocabulary and the ability to make use of the available linguistic context are however preserved with increasing age and can partly compensate for the age-related decreases in auditory functions (e.g., [24]). Sheldon et al. [25] examined the effect of age on the benefit from contextual information during listening. Participants had to repeat the sentence-final word of noise-vocoded sentences. This final word was either predictable from the sentence context (i.e., high-predictable sentences) or not. In some conditions, a prime was presented prior to the auditory sentence. The prime consisted of the undistorted auditory sentence with the target word replaced by a white-noise segment. For the high-predictable sentences, the age-related differences in the comprehension of the noise-vocoded speech disappeared in the prime-conditions. Sheldon et al. [25] suggest that these findings reflect the higher skill of elderly listeners at using supportive contextual information to compensate challenging listening conditions. This age-related increase in the ability to use contextual information is the result of more practice at using this information in daily life due to age-related reductions in auditory functioning [14, 25]. However, using these top-down processing skills to compensate for the distortions in the incoming auditory information requires effort, and when the available cognitive abilities and working memory capacity are insufficient to compensate for reduced auditory information, listening can become effortful [14, 26, 27, 28] and may result in fatigue.

Assessment of cognitive functions relevant in speech comprehension

Our department has recently developed a visual test to assess the ability of individuals to make use of linguistic context (i.e., verbal inference making; [29]). This Text Reception Threshold (TRT) test is a visual analogue of the widely-used Speech Reception Threshold (SRT) test [30] measuring speech comprehension in noise. The TRT is defined as the percentage of unmasked text required by the participant to read 50% of masked sentences without error. The sentences [30] are displayed one by one and the text is adaptively masked with vertical bars to estimate the TRT. At the start of each trial, the mask becomes visible and the text appears "behind" it in a word-by-word fashion. The timing of the appearance of each word is equal to the timing of each word in the original audio file. The preceding words remain on the screen until the sentence is completely presented, and then the sentence remains visible for 3.5 sec. Figure 1 illustrates the stimuli used in the Text Reception Test. The adaptive procedure applied in the TRT test is similar to the procedure applied in the SRT test [29]. The examination of the association between the performance on this visual test and speech comprehension in noise gives more insight into the factors explaining the interindividual variance in speech processing in noise.

Figure 1: Example of the stimuli used in the Text Reception Test [29].

Zekveld et al. [29] and George et al. [3] showed that the performance on the TRT test is significantly associated with the performance on the SRT test, thereby confirming the involvement of non-auditory linguistic skills in speech comprehension. In a group of hearing impaired participants [3], the TRT was significantly associated with the SRT in modulated noise, after controlling for temporal auditory acuity. In another study performed by our group [31], we examined the benefit obtained from partly incorrect textual output from an Automatic Speech Recognition (ASR) system during speech comprehension. Listeners with hearing impairment were asked to indicate the amount of benefit obtained from the supportive text during the comprehension of short stories presented in babble-noise. Participants with good TRTs reported higher speech comprehension support from the text than participants with worse TRTs.

As described by Kramer et al. [32], it would be interesting to further develop specific cognitive tests in order to provide more insight into the interindividual differences in speech comprehension ability. Currently, we are performing a pilot study in which a slightly adapted version of the TRT test is used in which the duration of the presentation of the sentences on the screen is shortened. This increased time-pressure during reading may increase the involvement of working memory, which may result in a stronger association of the TRT and speech comprehension.

Physiological measures of top-down speech processes

Increased reliance on top-down functions to increase speech comprehension can be examined with physiological measures. For example, functional magnetic resonance imaging (fMRI) can reveal to which extent the cognitive processes required by more difficult listening conditions result in increased activation of brain areas involved in these top-down functions. Additionally, using conjunction analyses, the overlap in brain regions activated by (slightly) different tasks or stimuli can reveal whether two conditions share a common function [33].

Speech comprehension is associated with activity in the anterior part of the superior temporal sulcus (STS), and the processing of complex auditory stimuli that depend on working memory processes activate areas in the left inferior frontal gyrus (IFG; [34, 35, 36, 37, 38]). Scott et al. [39] presented speech in noise stimuli at varying signal to noise ratios. They observed that more activation was evoked in the left inferior frontal cortex and the dorso-medial premotor cortex at the more difficult SNRs. This suggests that these regions are involved in compensatory cognitive processes that support speech comprehension when the listening task becomes harder. Davis and Johnsrude [40] also observed that prefrontal areas, including the premotor cortex and areas in the inferior frontal gyrus, exhibited an increased response to degraded stimuli compared to clear speech and an unintelligible baseline. They interpreted this finding as reflecting compensatory processes that apply at a nonacoustic level. Jobard et al. [41] examined the shared neural network during listening to speech and during reading. Bilateral areas were activated during both listening and reading tasks, including the posterior part of the left middle temporal gyrus, close to the superior temporal sulcus, extending in the occipito-temporal junction. Additional large activation clusters were observed in the left frontal cortex in the precentral gyrus and the pars opercularis and pars triangularis of the IFG. Irrespective of modality (auditory versus visual), the activity in areas in the left precentral gyrus and left IFG increased when the complexity of the verbal information increased (i.e., when comparing the activation evoked by words with that evoked by sentences and texts). Zekveld et al. [38] showed that increasing the SNR of speech presented in noise results in a sigmoid-shaped increase of activation as function of SNR in a frontal-temporal network of brain areas. At relatively high SNRs (i.e., at easier speech comprehension levels), the activation in the frontal brain areas was smaller compared to the activation in temporal brain areas, which is in line with the involvement of frontal areas in top-down processing of the speech and bottom-up processing of the speech in temporal brain areas.

The value of using brain imaging techniques besides behavioral measures of speech comprehension is that they can indicate to what relative degree the allocation of top-down processing resources have contributed to speech comprehension [42, 43, 44]. Specifically, individual differences in the amount of effortful processes required for successful speech comprehension will provide additional information that may not be revealed by speech intelligibility measures [33, 43]. In other words, examination of the brain areas associated with top-down processing of speech can indicate how individual listeners reach a certain speech comprehension performance. Research that examines the relation between relevant cognitive abilities and physiological measures will further contribute to the development of models of speech comprehension in difficult listening conditions.

Summary and Conclusion

In summary, top down cognitive functions are relevant for speech comprehension in noise, and interindividual differences in these functions can explain part of the variance in speech comprehension performance that is not explained by differences in auditory functioning. Future studies could further focus on the specific cognitive variables associated with speech comprehension in difficult conditions, and could lead to the development of clinically applicable tests that can be used to give more insight into the sources of communication difficulties on an individual basis. The application of brain imaging techniques in experimental settings can provide knowledge upon the neural basis of speech comprehension in adverse listening conditions, and upon the influence of cognitive abilities in the amount of processing capacity required for a speech comprehension task.

References

[1] Humes, L. E., & Christopherson, L. Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits. Journal of Speech and Hearing Research **34** (1991), 686–693.

[2] Pichora-Fuller, M. K., & Souza, P. E. Effects of aging on auditory processing of speech. International Journal of Audiology **42** (2003), S11-S16.

[3] George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., & Houtgast, T. Auditory and nonauditory factors affecting speech reception in noise by older listeners. The Journal of the Acoustical Society of America **121** (2007), 2362-2375.

[4] Houtgast, T & Festen, J. M. On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. International Journal of Audiology **47** (2008), 287-295.

[5] Plomp, R., & Mimpen, A. M. Speech-reception threshold for sentences as a function of age and noise level. The Journal of the Acoustical Society of America **66** (1979), 1333–1342.

[6] Bronkhorst, A. W., Brand, T., & Wagener, K. Evaluation of context effects in sentence recognition. Journal of the Acoustical Society of America **111** (2002), 2874-2886.

[7] Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., & Strauser, L. E. Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effects of presentation rate. Journal of Speech, Language, and Hearing Research **50** (2007), 283-303.

[8] Akeroyd, M. 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. International Journal of Audiology **47** (2008), S53-S71.

[9] Best, J.B. Cognitive Psychology. West Publishing Company, St. Paul, Minneapolis, 1995.

[10] Baddeley, A.D., & Hitch, G. Working memory. In: Bower, G.A. (ed). The psychology of learning and motivation. (1974). Academic Press, pp 48-79.

[11] Baddeley A. Is working memory still working? American Psychologist **56** (2001), 849–864.

[12] Daneman, M., & Carpenter, P. A. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior **19** (1980), 450-466.

[13] Lunner, T. Cognitive function in relation to hearing aid use. International Journal of Audiology **42** (2003), S49-S58.

[14] Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. How young and old adults listen to and remember speech in noise. The Journal of the Acoustical Society of America **97** (1995), 593-608.

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

[16] Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. Cognition counts: A working memory system for ease of language understanding. International Journal of Audiology **47** (2008), S99-S105

[17] George, E. L. J., Kramer, S. E., Festen, J. M., & Houtgast, T. Clinical relevance of auditory and nonauditory factors affecting speech reception in noise. Submitted for publication (2008) *Ear and Hearing*.

[18] Humes, L.E. Factors underlying the speech-recognition performance of elderly hearing-aid wearers. Journal of the Acoustical Society of America **112** (2002), 1112-1132.

[19] Pichora-Fuller, M. K. Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. International Journal of Audiology **42** (2003), S59-S67.

[20] Plomp, R. The Intelligent Ear. On the Nature of Sound Perception. Lawrence Erlbaum Associates, Mahwah, NJ, 2002.

[21] Suprenant, A. M., & Watson, C. S. Individual differences in the processing of speech and nonspeech

sounds by normal-hearing listeners. The Journal of the Acoustical Society of America **110** (2001), 2085-2095.

[22] Pichora-Fuller, M.K. Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. International Journal of Audiology **47** (2008), S72-S82.

[23] Van der Elst, W., Van Boxtel, M., Van Breukelen, G., Jolles, J. The letter digit substitution test: normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): Influence of age, education, and sex. Journal of Clinical and Experimental Neuropsychology **28** (2006) 998-1009.

[24] Wingfield, A. Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. Journal of the American Academy of Audiology **7** (1996), 175-182.

[25] Sheldon, S. Pichora-Fuller, M.K., Schneider, B.A. Priming and sentence context support listening to noise-vocoded speech by younger and older adults. Journal of the Acoustical Society of America 123 (2008), 489-499.

[26] Kramer, S.E., Kapteyn, T.S., Festen, J.M. & Kuik, J.D. Assessing aspect of hearing handicap by means of pupil dilatation. Audiology **36** (1997), 155-164.

[27] Wingfield A., & Tun, P.A. Cognitive supports and cognitive constraints on comprehension of spoken language. Journal of the American Academy of Audiology **18** (2007), 548-558.

[28] Wingfield, A., Tun, P.A., & McCoy, S.L. Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. Current Directions in Psychological Science **14** (2005), 144-148.

[29] Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T., & Houtgast, T. The development of the Text Reception Threshold test: A visual analogue of the Speech Reception threshold test. Journal of Speech, Language, and Hearing Research **50** (2007), 576-584.

[30] Plomp, R., & Mimpen, A. M. Improving the reliability of testing the speech reception threshold for sentences. Audiology **18** (1979), 43-52.

[31] Zekveld, A.A., Kramer, S.E., Kessens, J.M., Vlaming, M.S.M.G., & Houtgast, T. User evaluation of a communication system that automatically generates captions to improve telephone communication. Trends in Amplification **13** (2009), 44-68.

[32] Kramer, S.E., Zekveld, A.A., & Houtgast, T. Measuring cognitive factors in speech comprehension: The value of using the Text Reception Threshold test as a visual equivalent of the SRT test. Submitted for publication (2008), Scandinavian Journal of Psychology.

[33] Rudner, M., & Rönnberg, J. Towards a functional ontology for working memory for sign and speech. Cognitive Processing **7** (2006), S183-S186.

[34] Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., & Ward, B.D. Neural correlates of sensory and decision processes in auditory object identification. Nature Neuroscience **7** (2004), 295-301.

[35] Crinion, J.T., Lambon-Ralph, M.A., Warburton, E.A., Howard, D., Wise, R.S.J. Temporal lobe regions engaged during normal speech comprehension. Brain **126** (2003), 1193–1201.

[36] Müller, R.-A., Kleinhans, N., Courchesne, E. Broca's area and the discrimination of frequency transitions: A functional MRI study. Brain and Language **76** (2001), 70–76.

[37] Specht, K., Shah, N.J., Jäncke, L. Bilateral inferior frontal networks are involved in speech perception processes. NeuroImage **11** (2000), S292.

[38] Zekveld, A.A., Heslenfeld, D.J., Festen, J.M., & Schoonhoven, R. Top-down and bottom-up processes in speech comprehension. NeuroImage **32** (2006), 1826-1836.

[39] Scott, S.K., Rosen, S., Wickham, L., Wise, R.J.S. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. Journal of the Acoustical Society of America **115** (2004), 813– 821.

[40] Davis, M.H., Johnsrude, I.S. Hierarchical processing in spoken language comprehension. Journal of Neuroscience **23** (2003), 3423–3431.

[41] Jobard, G., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyera, N. Impact of modality and linguistic complexity during reading and listening tasks. NeuroImage **34** (2007), 784–800.

[42] Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., Kleinschmidt, A. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cerebral Cortex **14** (2004), 247–255.

[43] Pichora-Fuller, M. K. Audition and cognition: What audiologists need to know about listening. In C. Palmer, & R. Seewald (Eds.). Hearing care for adults (pp. 71-85). Phonak, Stäfa, Switzerland, 2006.

[44] Obleser, J., Wise, R.J.S., Dresner, M.A., & Scott, S.K. Functional integration across brain regions improves speech perception under adverse listening conditions. The Journal of Neuroscience **27** (2007), 2283-2289.