Influence of Vocabulary Knowledge & Lexical Access Times on Speech Intelligibility in Different Acoustic Conditions

Rebecca Carroll[1,2], Anna Warzybok[1,3], Esther Ruigendijk[1,2], Birger Kollmeier[1,3]

[1]Cluster of Excellence ‘Hearing4all’, Oldenburg

[2]Institute of Dutch Studies, University of Oldenburg

[3]Medizinische Physik, University of Oldenburg

Schlüsselwörter: Göttingen Sentence Test, reverberation, cognition, lexical access, vocabulary size

Introduction

Several researchers have suggested “linguistic skills” to correlate with speech intelligibility or recognition scores. Although linguistic “skills” or knowledge may refer to a multitude of language aspects, vocabulary size has often been considered a useful measure of linguistic skills or abilities (e.g., Benard et al, 2014; Besser et al, 2015; Conway et al, 2014). The more words a listener knows, the better his relative speech intelligibility scores. Benard et al (2014) investigated phoneme restoration by measuring speech intelligibility of Dutch everyday sentences (Versfeld et al, 2000) that were interrupted at rates of 1.25 and 2.50 Hz, with or without the addition of noise bursts in the silent gaps, and either at normal speech rate or a speech rate that was slowed by 50%. Speech intelligibility scores in percent correct were compared with a measure of receptive vocabulary knowledge (Dutch version of the Peabody Picture Vocabulary Test; Bell et al, 2001) and a Dutch version of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2012), respectively. They found receptive vocabulary, but not intelligence scores, to correlate with speech intelligibility by 12 listeners with normal hearing; the strongest correlation being one of PPVT and interrupted speech at the 2.50 Hz rate with silent gaps ($r = .851$; $p < .001$). Conway and colleagues (2014) showed that Stroop and Word tests for American children predicted the degree of facilitation by sentence context at the sentence final word in children with and without a cochlear implant. Working memory, as tested by the Digit Span, predicted accuracy of target word repetition at three sentence positions. Benichov and colleagues (2012) found, however, that vocabulary knowledge, as measured by a word naming test and word reading with unreliable phoneme-grapheme-correspondence, did not significantly contribute to word recognition performance in a group of 19–85 year old American listeners.

Vocabulary knowledge thus seems to vary in its predictiveness for speech recognition scores. This could be due to the language or group of listeners tested, the type of vocabulary and intelligibility test (or speech recognition scoring) used. It is also notable that even in those cases were “vocabulary measures” are found to be significant predictors of speech recognition performance, no coherent theoretical explanation with respect to speech recognition and language processing has been offered. Although researchers tend to distinguish word knowledge (as in long-term memory representations) and process-related aspects (e.g., lexical access or word naming times), it is not quite clear how these different aspects of the mental lexicon are related, or why/ how they may relate to different aspects of speech perception and word repetition. There also appears to be a lack of research on the role of vocabulary knowledge and lexical access for speech audiometric measures in the German language. In addition, different acoustic conditions may demand different aspects for correct speech recognition. For these reasons, we wanted to get a first impression of what lexical/ linguistic aspects may be relevant for speech recognition of a German speech audiometric test in different acoustic conditions. We therefore correlated speech recognition scores of sentences presented in different acoustic conditions with two measures of vocabulary size, a measure of lexical access times, and a measure of verbal working memory in young adults with normal hearing.

Material and Method

Participants. We tested 22 young listeners (18 – 34 yrs.; 15 women) with normal hearing (HL ≤ 15 dB between the seven octave frequencies, but one frequency (4 or 8 kHz) ≤ 20 dB HL). The group consisted of both academics and non-academics. Listeners were presented with the Göttingen sentence test (GÖSA; Kollmeier & Wesselkamp, 1997) in 6 acoustical conditions (see below), two standardized measures of vocabulary size, a lexical decision test, and the German Reading Span Test.

Acoustical Conditions. GÖSA test lists of 20 sentences each were presented (1) in standard test noise with fixed Signal-to-Noise ratios (SNR) of -4 and -6 dB, (2) interrupted at a 2.50 Hz rate and a 50% duty cycle (cf. Benard et al., 2014), (3) with reverberation times of 2.03 and 3.24 sec, and (4) a combination of 3.24 sec reverberation time and noise at 7 dB SNR. Measurements tested intelligibility in percent (%) correct repetitions. We used word scoring for all conditions. The parameters of conditions 1 and 3 were selected based on the study by Rennies et al (2015) and Warzybok et al (2015). Based on the predictions of the Speech Transmission Index (STI), intelligibility of about 50% is expected in a condition with SNR of -6 dB as well as in reverberation of 3.24 sec. In the conditions with SNR of -4 dB as well as in reverberation of 2.03 sec 80% intelligibility is expected. The combination of noise and reverberation should then result in intelligibility lower than 50% recognition rate. Seven participants did not listen to the reverberated speech and not to the -4 dB SNR condition. Statistical analyses thus only included 15 participants for these conditions.

Presentation and Apparatus. All GÖSA stimuli were presented diotically using the Oldenburg Measurement Application, over free-field equalized Sennheiser HDA200 headphones, amplified by either an Earbox 3.0 High Power, or an RME Fireface UCX. Participants were seated in a sound-attenuated booth. The measurement setup was calibrated to dB SPL using Brüel & Kjær instruments. PPVT words were presented auditorily with pre-recorded soundfiles over a Genelec loudspeaker, at 65 dB SPL.

Vocabulary Measures. The Wortschatztest (WST; Schmidt & Metzler, 1992) is a standardized pen-and-paper test of receptive vocabulary. Participants were presented 42 lines of five pseudo words and one existing word each, which they had to identify. Raw test scores were translated to the respective standardized z-scores. The WST is assumed to test recognition of the (orthographic) word form. Semantic knowledge is not required.

In the German version of the Peabody Picture Vocabulary Test (PPVT; Buhlheller & Häcker, 2003), participants were visually presented 4 pictures and an auditorily presented word. The task was to indicate the picture that best represented the target word (see Fig.1). The test consists of 89 trials with increasing picture-matching difficulty. To perform well on this test, individuals not only needed to be familiar with the (acoustic) word form but also have a detailed semantic representation of the target word to correctly distinguish the correct picture from its three semantically similar and/or related competitors.

The Lexical Decision Test (LDT) presented short 4-letter combinations on a computer screen. Forty items were monosyllabic pseudowords (i.e., non-existent words that are structurally possible but carry no meaning in German, e.g., MAND). Forty items were monosyllabic existing words, of which half (n=20) occur frequently and half occur infrequently in the language. Frequency of occurrence was established using the Leipzig Wortschatz corpus (wortschatz.uni-leipzig.de). The participants' task was to decide whether a given letter combination represented an existing German word. Responses were collected via button press. Log-transformed reaction times (RT) were calculated for correctly answered trials for pseudowords, high frequency, and low frequency words. Lexical access was defined as the relative RT, i.e., the difference between pseudo- and actual words (abbreviated here as $LDT_{\Delta PW-W}$). Frequent words are more likely to be pre-activated than less frequent words; RT are therefore bound to be much faster. On the flipside, (relative) RT to frequent words may not reflect a full lexical search as would be assumed for infrequent words. We thus opted to use the average of frequent and infrequent words for the difference between pseudo and actual words, but also tested the logRT for each condition separately.

Verbal Working Memory. We tested the German Reading Span Test (RST; Carroll et al, 2015) to determine whether speech intelligibility scores were correlated with verbal working memory. Such a correlation has been suggested by several researchers (e.g., Akeroyd, 2008; Rönnberg et al, 2013) and critically discussed (e.g., Besser et al, 2013). More importantly for us, we were interested in a correlation between RST, GÖSA, and vocabulary measures: If listeners with a larger receptive vocabulary (PPVT, WST) require more time for lexical access (positive correlation of PPVT/WST and lexical decision test), they would have to have a higher span score as well to perform well in GÖSA (interactive correlation of PPVT, LDT, GÖSA).

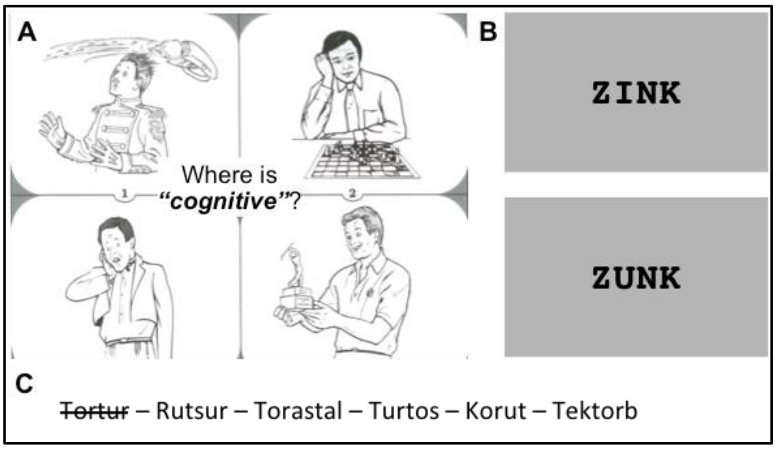

Figure 1: Examples of PPVT (panel A), LDT (panel B) with existing (top) and pseudo word (bottom), WST (panel C)

Results

A. GÖSA in different acoustical conditions

Table 1 summarizes the speech recognition scores (in percent correct) of GÖSA for the six acoustical conditions (a), as well as the scores for the vocabulary related tasks (b). Speech intelligibility scores for the original GÖSA test noise and for the reverberant conditions were close to the predictions at 50 % and 80%, respectively. The mean RST score was slightly better than the scores reported in Carroll et al (2015) for older listeners with hearing impairment (*M* = 21). Both WST and PPVT scores were in the top fourth on the respective scoring scales.

Table 1: (a) Mean GÖSA speech recognition performance in different acoustic conditions
(b) Mean scores for the linguistic measures and pure-tone average (PTA-4)

(a) Condition	N	Mean (± SD)	STI prediction
SNR-6	22	50.9 ± 16.1	50 %
SNR-4	15	85.1 ± 10.9	80 %
SNR7_T60_3.25	22	35.6 ± 11.7	< 50 %
T60_3.24	15	49.7 ± 15.0	50 %
T60_2.03	15	81.7 ± 9.0	80 %
Interrupted	22	58.6 ± 11.4	n/a

(b) Additional Measure	N	M	SD
WST	22	0.45	0.71
PPVT	22	80.41	17.43
$LDT_{\Delta PW\text{-}W}$	22	0.09	0.05
RSTcorr	15	24.62	6.78
PTA-4	22	3.34	3.03

B. Correlation of Speech Recognition Performance with Vocabulary Measures

Initial correlations of speech recognition performance in the different acoustic conditions with the different vocabulary and cognitive measures showed an obscure pattern (cf. Table 2). All measures showed at least one significant correlation to one of the GÖSA conditions. But not one measure seemed to correlate with all acoustic conditions, and correlations were not very high. Interrupted speech and reverberated speech (3.24 sec) did not correlate with any of the measures under investigation.

Table 2: Correlations of speech recognition scores and vocabulary measures.

Acoustic Condition	N	WST r	WST p	PPVT r	PPVT p	RST r	RST p	$LDT_{\Delta PW\text{-}W}$ r	$LDT_{\Delta PW\text{-}W}$ p
SNR-6	22	.263	*n.s.*	.245	*n.s.*	**.375**	**.042***	**.575**	**.002***
SNR-4	15	**.605**	**.008***	**.650**	**.004***	**.643**	**.005***	.067	*n.s.*
SNR7/T60_3.25	22	.133	*n.s.*	.159	*n.s.*	.269	*n.s.*	**.660**	**.002***
T60_3.24	15	-.138	*n.s.*	- .052	*n.s.*	.408	*n.s.*	.232	*n.s.*
T60_2.03	15	**.508**	**.027***	.407	*n.s.*	.390	*n.s.*	- .091	*n.s.*
Interrupted	22	- .055	*n.s.*	- .027	*n.s.*	.198	*n.s.*	.111	*n.s.*

To account for multiple comparisons, we calculated stepwise linear regression models. First one model over all conditions, followed by one model per acoustic condition. We considered age, hearing status, WST, PPVT, LDT (log RT of pseudo words (PW), high frequency (HF) and low frequency (LF) words, and RT differences of PW and actual words ($_{\Delta PW\text{-}W}$), respectively), and RST scores as possible predictors. Predictors entered the model if the *F* value changed significantly at the 0.05 level. Table 3 lists the summaries of the regression models. The overall model returned PPVT as the only significant predictor (see Table 3), although the effect size was rather low. In the models per acoustic condition, WST seemed to be related to GÖSA scores more than PPVT. The $LDT_{\Delta PW\text{-}W}$ only correlated with SNR7/T60_3.24. None of the intelligibility scores were significantly correlated with hearing status (mean PTA of 0.5, 1, 2, 4 kHz).

There was a significant negative correlation of vocabulary size (PPVT) and mistakes in the LDT (r = - .73), suggesting that the more words a listener knew, the fewer mistakes he would make during lexical decision. Moreover, age correlated significantly with vocabulary size (WST: r = .64; PPVT: r = .60). Vocabulary size also correlated with verbal working memory (WST vs. RST: r = .64). Verbal working memory (RST) *per se* could not be reliably attributed to the performance variance in any condition (in the regression models, Table 3), although single correlations (cf. Table 2) may have suggested this, at least for the two noise conditions.

Table 3: Regression models per acoustic condition.

Acoustic Condition	Regression Model	N	r	r^2	Corrected r^2	Change in F	*p*
Overall	PPVT	22	.244	.059	.049	5.75	.019*
SNR -6	LDT_{RT_PW}	22	.586	.347	.301	7.45	.016*
SNR -4	WST	15	.650	.423	.379	9.54	.009*
SNR7_T60_3.25	$LDT_{\Delta PW-W}$	22	.660	.435	.395	10.80	.005*
T60_3.24	Age	15	.569	.324	.272	6.23	.027*
	Age, WST	15	.748	.559	.486	6.41	.026*
	Age, WST, PPVT	15	.838	.702	.620	5.24	.043*
T60_2.03	--	15	--	--	--	--	--
Interrupted	--	22	--	--	--	--	--

Discussion

We tested normal hearing listeners' word recognition of the Göttingen Sentence Test in different acoustical conditions. Their scores were compared to linguistic measures of vocabulary size, lexical access efficiency, and verbal working memory. Results showed a variable picture. Three observations are noteworthy:

1. *Speech intelligibility varied across listening conditions.* Percent correct scores were close to the predicted scores. PTA did not correlate with speech recognition scores. This suggests that the variance observed in the GÖSA conditions could not be explained by hearing status. Assuming linguistic knowledge and/ or cognitive measures to explain parts of the variance thus seemed reasonable. Given that vocabulary size may increase with age, we included this as a factor in the regression models.

2. *Not all linguistic measures appear to be correlated to the same degree*. Vocabulary size as measured by the standardized WST correlated with speech recognition scores when presented in the original test noise at -4 dB (close to 80% intelligibility), but not at a lower SNR of -6 dB (close to 50 % intelligibility); at the lower recognition rate of reverberation (3.24 sec reverberation time) only in combination with age and PPVT, but not in any other acoustical condition. Lexical access as defined by the relative RT ($_{\Delta PW-W}$) in the LDT only correlated significantly in the combinatory acoustic condition of noise plus reverberation. This could be interpreted as lexical access efficiency being most strongly related to the very difficult acoustic listening condition (35.6 % speech recognition). The fact that the LDT measure was a significant predictor for the -6 dB SNR in the individual correlations would support such an interpretation. The fact that GÖSA scores in noise at an approximate recognition rate of 50 % (SNR-6) only seemed to be related to the log RT for pseudo words is difficult to interpret. It may simply be an underpowered effect or possibly an effect of general processing speed. Most notably, we could not replicate the Dutch findings of Benard et al (2014), who observed a very high correlation of speech recognition performance and PPVT scores. It has to be noted, however, that their correlation seemed to be due to one or two listeners only. Their participants also showed a much larger age range (21-63 yrs). It is not clear whether the low-performers in the Dutch data correspond to older participants.

3. *Correlations between linguistic measures suggest partial overlap of linguistic and/or cognitive processes* that may contribute to speech recognition scores. For example, correlations between WST, PPVT and RST suggest that RST may indirectly play a role in speech recognition scores, but not necessarily as directly as previously proposed (e.g., Rönnberg et al, 2013; Besser et al, 2013, 2015).

It is also reasonable to find age to be strongly related to vocabulary size, as (most of) our participants continuously broaden their knowledge, especially with respect to loanwords and technical terms.

It is thus reasonable to conclude that A) different acoustic listening conditions may relate to different linguistic and/or cognitive aspects, B) individual listeners may rely on different factors for speech recognition. Both conclusions have to be taken with caution: More data, especially in populations with greater inter-individual variation, is needed to substantiate any claims of linguistic factors playing a role in speech recognition. Our data suggest that C), any study trying to determine the role of "linguistic abilities" for speech recognition should consider more than just one measure, and possibly consider different linguistic aspects of speech processing.

References

Akeroyd, M (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*(1), S53-S71.

Bell, NL, Lassiter, KS, Matthews, TD, & Hutchinson, MB (2001). Comparison of the Peabody Picture Vocabulary Test–3rd Edition and Wechsler Adult Intelligence Scale–3rd Edition with university students. *J Clin Psychol, 57*, 417–422.

Benard, M, Mensink, JS, & Başkent, D (2014). Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. *J Acoust Soc Am, 135*(2), EL1-EL8.

Benichov, J, Cox, CC, Tun, PA, & Wingfield, A (2012). Word recognition within a linguistic context: Effects of age, hearing acuity, verbal ability, and cognitive function. *Ear Hear, 33*, 250-256.

Besser, J, Festen, JM, Goverts, ST, Kramer, SE, & Pichora-Fuller, MK (2015). Speech-in-speech listening on the LiSN-S Test by older adults with good audiograms depends on cognition and hearing acuity at high frequencies. *Ear Hear, 36*(1), 24-41.

Besser, J, Koelewijn, T, Zekveld, AA, Kramer, S, & Festen, J (2013). How linguistic closure and verbal working memory relate to speech recognition in noise: A review. *Trends Amplif, 17*(2), 75-93.

Brand, T, & Hohmann, V (2002). An adaptive procedure for categorical loudness scaling. *J Acoust Soc Am*, 112, 1597.

Buhlheller, S, & Häcker, HO (2003). *Deutschsprachige Fassung des PPVT-III für Jugendliche und Erwachsene.* Frankfurt: Swets.

Carroll, R, Meis, M, Schulte, M, et al. (2015). Development of a German reading span test with dual task design for application in cognitive hearing research. *Int J Audiol, 135,* 136-141.

Conway, CM, Deocampo, JA, Walk, AM, Anaya, EM, & Pisoni, DB (2014). Deaf children with cochlear implants do not appear to use sentence context to help recognize spoken words. *J Sp Lang Hear Res, 57*, 2174-2190.

Kollmeier, B, & Wesselkamp, T (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J Acoust Soc Am, 102*(4), 2412-2421.

Rennies, J, Schepker, H, Holube, I, & Kollmeier, B (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *J Acoust Soc Am*, 136(5), 2642-2653.

Rönnberg, J, Lunner, T, Zekveld, AA, 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.

Schmidt, K-H, & Metzler, P (1992). *Wortschatztest.* Weinheim: Beltz.

Versfeld, NJ, Daalder, L, Festen, JM, & Houtgast, T (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am, 107*, 1671–1684.

Warzybok, A, Rennies, J, Brand, T, Kollmeier, B (2015). Measurement and prediction of speech intelligibility in noise and reverberation for different sentence materials. *7th Speech in Noise Workshop,* Copenhagen, 8th-9th January, 2015.

Wechsler, D (2012). *Wechsler Adult Intelligence Scale*, 4th ed. (WAIS-IV-NL) Dutch version.