Bernstein, J. G. W., and Brungart, D. S. (**2011**). "Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio" J. Acoust. Soc. Am. **130**, 473-488.

Ching, T. Y. C., Dillon, H., Katsch, R., and Byrne, D. (**2001**). "Maximizing effective audibility in hearing aid fitting" Ear Hear. **22**, 212-224.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing" J. Acoust. Soc. Am. **88**, 1725-1736.

Fousek, P., Svojanovsky, P., Grezl, F., and Hermansky, H. (**2004**). "New nonsense syllables database — analyses and preliminary ASR experiments" Proceedings ICSLP 2004, Jeju, South Korea.

Gnansia, D., Pean, V., Meyer, B., and Lorenzi, C. (**2009**). "Effects of spectral smearing and temporal fine structure degradation on speech masking release" J. Acoust. Soc. Am. **125**, 4023-4033.

Hopkins, K., and Moore, B. C. J. (**2009**). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise" J. Acoust. Soc. Am. **125**, 442-446.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech" J. Acoust. Soc. Am. **123**, 1140-1153.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167-173.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials" J. Exp. Psych. **41**, 329-335.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference" J. Acoust. Soc. Am. **125**, 457-468.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired" J. Speech Hear. Res. **29**, 146-154.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers" J. Acoust. Soc. Am. **114**, 446-454.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshhold in fluctuating noise" J. Acoust. Soc. Am. **120**, 3988-3997.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues" Science **270**, 303-304.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Effect of spectral envelope smearing on speech reception. II" J. Acoust. Soc. Am. **93**, 1547-1552.

Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (**2009**). "The interaction of vocal characteristics and audibility in the recognition of concurrent syllables" J. Acoust. Soc. Am. **125**, 1114-1124.

Recognition rates and linguistic processing: Do we need new measures of speech perception?

THOMAS BRAND, VERENA USLAR, DOROTHEA WENDT, BIRGER KOLLMEIER

Medizinische Physik, Universität Oldenburg, D-26111 Oldenburg, Germany

Speech perception goes far beyond the recognition of phonemes, words, and sentences. The Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) were developed for investigating the interactions between the listener's linguistic and auditory capabilities in speech perception. Using these sentences with normal-hearing and hearing-impaired listeners in different listening conditions, a small but significant influence of the sentences linguistic complexity was detected. To some degree this influence was related to other cognitive measures of the listeners. In an eye-tracking experiment delayed eye movements for more complex sentences indicated a higher cognitive load during the speech recognition process. These delayed eye movements were sensitive even in conditions where classical recognition rate and speech reception measures were not sensitive because the recognition rate was near 100%.

INTRODUCTION

When speech perception is assessed in audiology in many cases recognition rates are measured. For this purpose different speech intelligibility tests are available in different languages, such as phoneme or word recognition tests. However, a principal problem of these tests is that they are not very efficient, because for a given number of tested speech items the confidence interval of the recognition rate estimate is relatively wide, due to the binomial distributed data. When the speech reception threshold (SRT) is measured using a test with a steep level-intelligibility curve, such as a sentence test, relatively precise measurements with a standard error of about one dB are possible within only a few minutes per listener. Such SRT tests work very well in many conditions and are a working horse in many applications. However, the SRT is usually located in a level range which is too low for some applications. For instance, some noise reduction algorithms used in hearing aids do only work properly at positive SNRs. The use of threshold values related to recognition rates higher than 50% (for example 80%) can sometimes help here because the signal-to-noise ratio is a few dB higher.

In this article we address the question if there might be further measures besides speech recognition rates and SRTs which might be useful in speech audiometry and which work even in conditions with a recognition rate near 100%. For this purpose we investigated how the linguistic complexity of test sentences influences recognition rates and SRTs as well as eye movements. One of our hypotheses is that an increased linguistic complexity of the used sentence material causes an increased cognitive load of the listener which results in increased SRTs if an interfering noise

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

is used. An uncontrolled variance of linguistic complexity of the used speech material may increase the variance of the SRT estimate. On the other hand a controlled variance of the linguistic complexity can be used to investigate the listener's capability to understand speech in different situations and to quantify his or her linguistic competence in contrast to his or her auditory capabilities.

A further hypothesis of this article is that delayed eye movements as measured using an eye tracking device can reveal an increase of the listener's processing load when listening to sentences with different linguistic complexity and that this is possible even though the recognition rate (nearly) 100%. In other words, that eye movements can be used as a measure for speech perception in conditions where recognition rates and SRTs can not provide any information.

This article reports on three studies. In the first study a possible parasitic effect of linguistic complexity on standard speech recognition tests is investigated. The second study describes the development of a new speech corpus with audiological and linguistically controlled sentences and how these sentences are perceived in different noise conditions. The third study investigates how eye movements can be used to track the perception of sentences with different degrees of linguistic complexity.

STANDARD AUDIOMETRIC SENTENCE TEST

The question that was addressed in a study by Uslar *et al.* (2011) was if there is a potential influence of linguistic complexity on a standard audiometric speech recognition test. For that purpose the Göttingen sentence test (Kollmeier and Wesselkamp, 1997) which is frequently used for speech audiometry in Germany was used. This test comprises balanced test lists with respect to intelligibility, phoneme distribution, and number of words. However, a close reanalysis showed that the sentences differ in linguistic complexity. The hypothesis was that these differences cause differences in intelligibility that may have been overseen in the optimization and evaluation of the test. In order to test this hypothesis, three new sentence lists were reassembled from a subset of the original material with different degrees of linguistic complexity. The first new sentence list (SIMPLE) contained linguistically simple sentences. The second sentence list (COMPLEX_V) contained sentences with verbs with more than one possible argument structure. The third sentence list (COMPLEX_S) contained sentences with non-canonical word order.

Twenty younger (26 ± 3 years) ontologically-normal listeners, eleven older (45 ± 4 years) otologically-normal listeners, and eleven aged matched (49 ± 5 years) listeners with hearing loss took part in the study. Listeners with hearing impairment had a mild to moderate, mostly sloping, sensorineural, postlingual hearing loss with no frequency worse than 80 dB HL. For each listener the SRT in speech shaped non-fluctuating noise was determined for each list.

Figure 1 shows the mean difference between SRTs of the complex and simple test lists. Younger listeners with normal hearing showed significantly worse SRTs on the complex lists than on the simple list. However, this difference could not be found for either of the older groups. Over all the effect introduced by the differences in linguistic complexity is relatively small – if present at all - even though sentences with the theoretically strongest linguistic effects have been selected for the COMPLEX_V and COMPLEX_S lists.

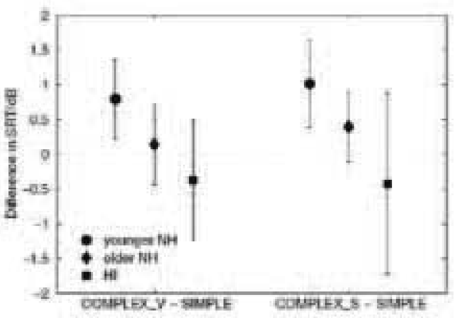

Fig. 1: Mean and standard deviation of the individual differences between SRTs in noise of COMPLEX_V and SIMPLE (left side) and COMPLEX_S and SIMPLE (right side) separated by groups of listeners. Positive differences indicate that the SIMPLE list was perceived better than the respective COMPLEX list. (From Uslar *et al.*, 2011)

Taken together we come to the following conclusions: Large differences of linguistic complexity between sentence lists should be avoided in the development of audiometric speech recognition tests in order to reduce uncontrolled variance. Fortunately, the variance of linguistic complexity in the original test lists of the Göttingen sentence test does not cause a significant variance in recognition rates and we assume that this also holds for other, similar speech recognition tests. On the other hand speech recognition tests that control linguistic complexity in a wide range may have the potential of revealing interactions between peripheral auditory processing and central language processing capabilities.

LINGUISTICALLY AND AUDIOLOGICALLY CONTROLLED SENTENCES (OLACS)

Development of OLACS

In order to get a well controlled tool for investigating linguistic effects in different listeners and hearing situations the Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) were developed. The sentences were controlled linguistically with respect to word order, embedding of relative clauses, ambiguity, and plausibility and provide the following types of sentences:

1. Transitive main clauses with canonical subject-verb-object word order and unambiguous allocation of grammatical functions and semantic roles (SVO).
2. Transitive main clauses with object-first word order and unambiguous allocation of grammatical function and semantic role (OVS).

3. Transitive main clauses with object-first word order and ambiguous case marking on the object noun phrase (ambOVS).
4. Intransitive main clauses with an embedded subject relative clause (relative pronoun is subject of the embedded clause) and unambiguous allocation of grammatical function and semantic role within the relative clause (SR).
5. Intransitive main clauses with an embedded object relative clause (relative pronoun is object of the embedded clause) and unambiguous allocation of grammatical function and semantic role within the relative clause (OR).
6. Intransitive main clauses with an embedded subject relative clause and initial function/role ambiguity (ambSR) (not used in this article).
7. Intransitive main clauses with an embedded object relative clause and initial function/role ambiguity (ambOR) (not used in this article).

In order to control the sentences audiologically, the sentence specific intelligibility for the whole material was assessed using 36 listeners with normal hearing.

In a first step, recognition rates were measured for fragments of 681 sentences which resulted from cutting each sentence into three pieces. SVO and OVS sentences were cut directly before and after the verb; relative clauses were cut at both commas. By this syntactical and context effects were reduced. Sentences with one or more words which were never understood were discarded to make the material more homogeneous.

In a second step, the intelligibility of the remaining 560 sentences was measured at an SNR of -7 dB using complete sentences. Sentences deviating by more than two standard deviations from the mean of the respective sentence type were discarded.

In a third step, discrimination functions for the remaining 360 sentences were determined and the final set of 40 sentences for each sentence type was selected.

Influence of linguistic complexity on intelligibility

The influence of the different sentence types on speech intelligibility in different noise conditions was investigated using 20 listeners with normal hearing and 17 listeners with sensorineural mild to moderate, mostly moderately sloping hearing loss (in the main frequency areas of speech hearing loss ranged between 20 and 50 dB HL and no value exceeded 80 dB HL). For each listener the SRT80/SRTN80 (speech level or SNR related to 80% recognized words) was measured using an adaptive procedure (Brand and Kollmeier, 2002).

Figure 2 shows the mean SRT80/SRTN80 for each sentence type in each listening condition (quiet, stationary noise, and fluctuating noise). Normal-hearing listeners are shown in the left panel. Hearing-impaired listeners are shown in the right panel. Overall, for normal-hearing listeners there are only small differences for the different sentence types. SR sentences produce the lowest (best) SRT80/SRTN80 in each listening condition. For sentences with the verb at the second position the ambOVS type produces the lowest SRT80/SRTN80 in quiet and in stationary noise, whereas in fluctuating noise the ambOVS sentences produce the highest SRTN80.

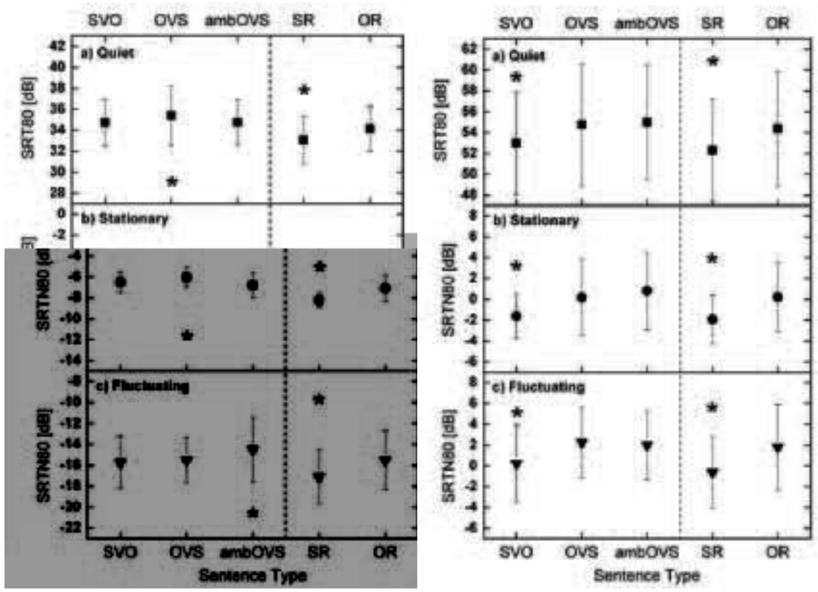

Fig. 2: mean SRT80/SRTN80 and standard deviation for each sentence type and each noise condition. Left panel: normal-hearing listeners, right panel hearing-impaired listeners. A black asterisk denotes significant differences (t-test, p < .05).

For listeners with hearing impairment the differences between the simple sentence types and the more complex ones are more pronounced, with significant differences ranging between 2 and 3 dB and for some listeners even exceeding 5 dB. There are no significant differences between listening conditions. Note that for hearing-impaired listeners the ambOVS condition always produces a mean SRT80/SRTN80 which is in the same range as for the OVS condition.

Cognitive measures

In order to test the hypothesis that the listeners' individual linguistic capabilities are correlated to other cognitive measures the listeners performed three cognitive tests:

1. The Stroop task which quantifies the subject's susceptibility of interference and general attention to a given task (Kim, Kim and Chun, 2005).
2. The Word-Span (forward) which quantifies the subject's verbal memory (span) capacity (Tewes, 1991).
3. The Digit-Span (backward) which quantifies the subject's memory and processing capacity (Tewes, 1991).

Fig. 3 shows the result of a principal component analysis (PCA) with the results of the three cognitive tests (Word Span, Digit Span and Stroop test), the mean hearing threshold (PTA, mean of the values of the audiogram at 0.5, 1, 2 and 4 kHz), age of the participant, and the SRT80/SRTN80 for all sentence types in each listening condition. The left panel shows the results of the normal-hearing listeners. The right panel shows the result of the hearing-impaired listeners. For normal-hearing listeners, the principal component analysis was conducted with orthogonal rotation (varimax), and the Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis (KMO = .83). The analysis revealed two clearly distinguishable main components (after analysis of the scree plot) which together explain 47 % of the variance in the data. Component 1 explains 26 % of the variance and mainly contains the SRT80/SRTN80 and the PTA. Component 2 explains 21 % of the variance and mainly contains the SRT80/SRTN80 and each individual and cognitive parameter entered into the analysis.

For hearing-imparied listeners, the Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis (KMO = .98). Again, the analysis revealed two clearly distinguishable main components (after analysis of the scree plot) which together explain 70 % of the variance in the data. Component 1 explains 60 % of the variance and mainly contains all speech intelligibility measures and the PTA. Component 2 explains 10 % of the variance and mainly contains the Word Span and Stroop test.

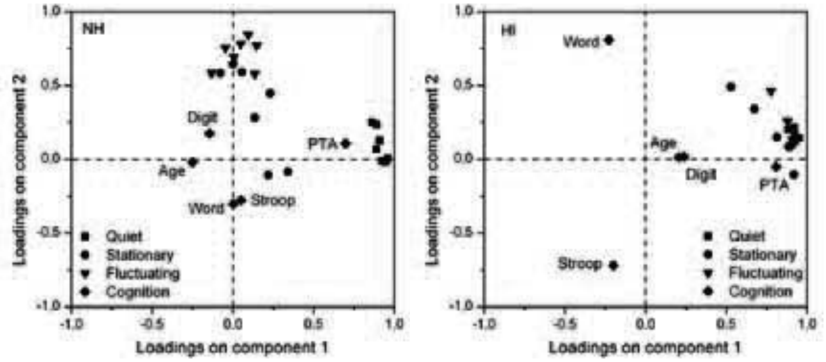

Fig. 3: loading diagrams of all variables included in the principal component analysis. Plotted are the loadings for each variable on component 1 against the loadings on component 2. Left panel: normal-hearing listeners; right panel: hearing-impaired listeners.

In Fig. 3 the loadings of each variable for component 1 is plotted against the loading of the respective variable for component 2. Loadings above 0.2 may be considered relevant. For normal- hearing listeners, all measurements in quiet and the PTA have high loadings on component 1 but small loadings on component 2. The measurements in noise (and here especially the measurements in fluctuating noise) and all individual and cognitive variables have relatively high loadings on component 2 but little influence on component 1. The age of the participants (age span about 10 years) is of little importance in both components. For hearing-impaired listeners, all speech measures and the PTA load highly on component 1. Additionally, Digit Span and age have a small positive influence on this component, whereas Word Span and Stroop test show small negative loadings for component 1 but high complimentary loadings on component 2. Also, some of the speech measures in noise have a relatively high loading on the more cognitive component 2.

The conclusions from this study using the OLACS material are:

Linguistic complexity in sentences with low semantic predictability has a small but significant effect on speech recognition for young normal-hearing listeners. In fluctuating noise it seems to be more pronounced, indicating greater cognitive load.

Hearing-impaired listeners show a more pronounced effect of syntactical complexity and problems with the ambiguous OVS sentences, which might be due to a higher preference for the subject-first order, which is typical for the German language.

In quiet the SRT80 for normal-hearing listeners is mainly correlated with the hearing threshold – perhaps a bit surprising for this group of listeners. In noise the SRTN80 correlates with the results of the Word Span and Stroop test, showing the importance of working memory and attention in adverse listening conditions. For hearing-impaired listeners the PTA clearly governs the results of speech intelligibility measures, with age and memory (through the Digit Span) influencing the results in a small bit significant way.

Taken together with the results of the study described above using the Göttingen sentence test (Uslar *et al.*, 2011), the results indicate, that hearing-impaired listeners generally seem to operate on higher cognitive load, and seem to employ a different listening strategy than young normal-hearing listeners, namely a strong subject-first preference and heavy reliance on semantic content. This strategy seems to benefit hearing-impaired listeners in everyday sentences with relatively high predictability.

EYE MOVEMENTS DURING RECOGNIZING SENTENCES WITH DIFFERENT LEVELS OF LINGUISTIC COMPLEXITY

In the next study (Wendt *et al.*, *in preparation*) the OLACS corpus and corresponding graphical material were used to investigate syntax-related influences on eye movements during sentence comprehension in 17 normal-hearing participants. A novel data analysis method was developed calculating the Target Detection Amplitude (TDA) which describes the tendency to fixate the target during sentence recognition.

Method

For each sentence a picture set consisting of two pictures was drawn. The two pictures show the same scene with roles of agent and patient interchanged. Each picture set is divided into three regions of interest (ROI): target picture (ROI 1), competitor picture with interchanged roles (ROI 2) and background (ROI 3). An example picture set is shown in Figure 4. In a visual world paradigm (Tanenhaus *et al.*, 1995) the OLACS picture set was presented visually on a computer screen concurrently with a spoken sentence of OLACS via headphones.

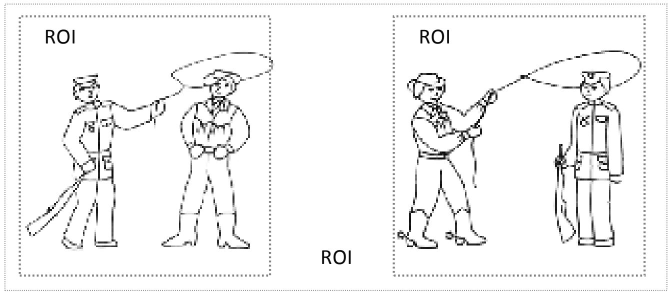

Fig. 4: Example picture set of the sentence “Der gute Soldat fängt den frechen Cowboy.” (Engl: “The good soldier catches the cheeky cowboy”). A picture set consists of two single pictures. The dashed lines indicate the three regions of interest (ROI) and are not visible for the participants.

The fixation rate of the background (ROI 3) is not considered, consequently the fixation rates of ROI 1 and ROI 2 do not add to 100%. Only trials in which the participants selected the correct picture are considered in order to analyze eye movements that reflect the correct recognition process.

	Seg 1	Seg 2		Seg 3	Seg 4		Seg 5		Seg 6
SVO	-	**DER**	liebe	Drache	fesselt	den	großen	Panda.	Response time
		the	nice	dragon$_{MAS}$	ties up	the	Big	Panda	
OVS	-	**DEN**	großen	Panda	Fesselt	der	liebe	Drache.	Response time
		the	big	Panda	ties up	the	nice	dragon	
Amb OVS	-	Die	liebe	Prinzessin	Fängt	**DER**	schnelle	Dieb.	Response time
		the	nice	princess	catches	the	fast	Thief	

Table 1: Time segments for different sentence types. Capital letters: Point of disambiguation

Since sentences differ in length, a time alignment was employed in order to enable comparisons across sentences. This was achieved by dividing each sentence into six segments as shown in Table 1. To synchronize the segment boarders, the first five segments were individually aligned to a fixed length of 100 samples using interpolation and resampling. The length of the Segment 6 depended on the participant’s reaction time and was resampled to a maximum length of 200 samples. For reaction times of more than 2000 ms, the signal was cut to a maximum length of 200 samples.

Calculation of the target detection amplitude (TDA)

The eye tracking data is transformed into a bias corrected Target Detection Amplitude (TDA) which quantifies the listener’s tendency to fixate the target. The TDA is computed in three steps:

In the first step, target and competitor fixation rates as functions of time are calculated on sentence basis distinguishing whether the target picture was shown on the left or on the right side. This distinction is necessary because, in general, higher fixation rates for pictures presented on the left side are observed, independently of the position of the target picture. This bias, which is especially noticeable for Segment 1, is assumed to be due to the reading direction and due to the fact that the agent is always presented on the left side of each picture.

In the second step the bias is compensated as follows: The target-position-dependent mean fixation rates of the competitors (for the left and the right side, respectively) are subtracted from the target-position-dependent fixation rates for the target. As a result two bias corrected fixation rates are obtained for the target on the left and the right side, respectively, which are subsequently added to get the bias corrected TDA. The *TDA* was computed for every sentence type and every participant.

In the third step a post processing of the TDA is performed: First, a Gaussian filter with a kernel size of 35 samples is used to smooth the TDA. And finally, the 95% confidence interval for the mean TDA over all participants is calculated using the bootstrap resampling procedure (see Efron and Tibshirani, 1993).

Results and discussion

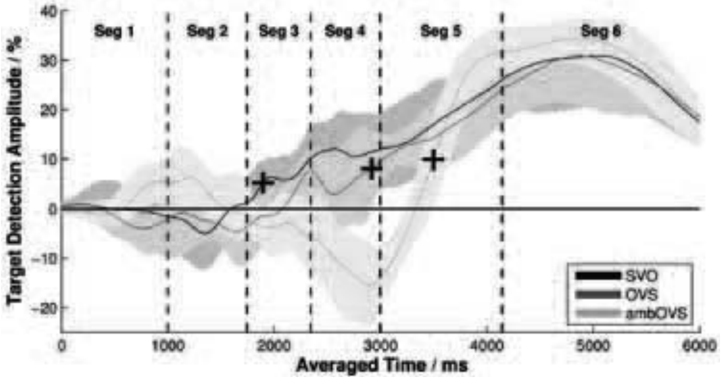

Fig. 5: Mean Target Detection Amplitude (TDA) of different sentence types in quiet. The plus signs indicate decision moments. Dashed vertical lines show averaged segment borders (see Table 1).

The mean TDAs across all listeners for three different sentence types are shown in Figure 5 (quiet) and Figure 6 (fluctuating speech shaped noise). Plus symbols denote the *decision moment* defined at the point in time at which the 95% confident interval of the TDA deviates above zero for the first time. The dashed vertical lines reflect the averaged segment borders as shown in Table 1.

For the unambiguous SVO in quiet (Figure 5), the decision moment is observed at 1900 ms, i.e. shortly after the first noun phrase (“der”) and during the agent of the spoken sentences is presented (cf. Segment 3 in Figure 5). Since the early case-marking of the first noun phrase (“der”) in this condition caused an early thematic role assignment, listeners were able to identify the noun-phrase referent as the agent of the depicted event and to recognize the target very early in the spoken sentence. Thus, the time patterns of the TDA correlate with the case-marked mediation of the relevant depicted event.

Just like the SVO, the first noun phrase (“den”) of the unambiguous OVS condition contains information about the role relations at the beginning of the spoken sentence. Despite this early point of disambiguation, the decision moment occurs about 2920 ms after stimulus onset, i.e. during Segment 4.

Although the confident intervals of the SVO and OVS sentence types do overlap in all segments, the decision moments differs by 1000 ms between both unambiguous sentence types. The non-canonical sentence structure of the OVS condition interferes with the subject-first preference of the listeners and probably requires additional processing costs for the working memory load (Schriefers *et al*., 1995). Thus, more processing difficulties for the OVS condition than for the SVO condition are expected and, therefore, a delayed point in time at which the listeners can disambiguate between target and competitor.

A different time course is observed regarding the object-first sentences with a late point of disambiguation (ambOVS). The decision moment for the ambOVS sentence type is observed at 3500 ms, i.e. more than 500 ms delayed in time compared to the decision moment of the unambiguous OVS sentences. A delayed increase of the TDA is expected since the verb of the embedded sentence disambiguated the ambOVS sentences. Hence, the ambiguity causes a delayed recognition of the non-canonical word order.

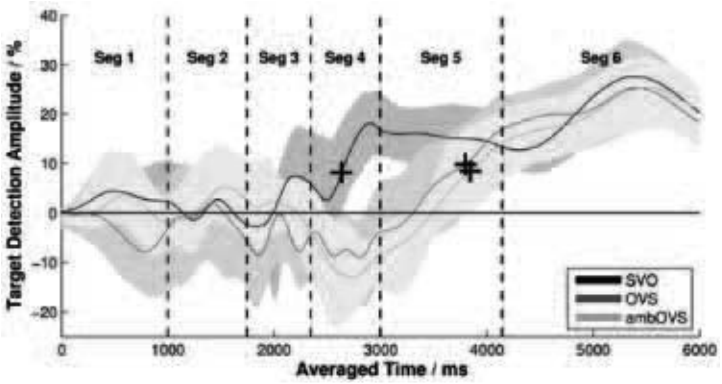

Fig. 6: Mean Target Detection Amplitude (TDA) of different sentence types in fluctuating speech shaped noise.

Negatives TDAs, which are observed during Segment 4 and Segment 5 for both object-first sentence types (OVS and ambOVS), reflect significant more fixations on the competitor picture due to the interpretation as being a subject-first structure at the given time interval. Hence, the spoken sentence was misinterpreted by the listeners at these points in time. The decrease of the TDA coincides with the point in time at which the verb was mentioned. This is an effect of canonicity (Knoeferle *et al.*, 2008). Up to the point in time at which the sentence was disambiguated, listeners are expected to choose the most preferred identity for the first noun phrase that is mostly the agent of the sentence. In case of an ambiguous object-first sentence listeners have to validate their first interpretation when the point of disambiguation is reached and they realize that their first role assignment of the object and subject is wrong.

Different TDAs can be observed for fluctuating speech shaped noise (Figure 6): All three sentence types exhibit delayed decision moments compared to quiet. For SVO sentences the delay is about 700 ms. The difference between decision moments of OVS and ambOVS sentences found in quiet disappears in noise. Furthermore, negative TDAs occur for unambiguous OVS sentences in Segment 4 indicating that listeners misinterpret the sentence as an SVO sentence at this point in time.

Taken together, we can conclude: The decision moment is influenced by sentence type and noise. Negative TDAs occur for object-first sentence types which are syntactically more complex than subject-first sentences. The observed negative TDAs are temporally located before the point of disambiguation of the spoken sentence and suggest a misinterpretation of sentence by the listeners.

GENERAL CONCLUSIONS

- In the Göttingen sentence test linguistic complexity does not influence recognition rates of complete test lists. We assume that this also holds for other sentence intelligibility tests.
- The Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) use linguistic complexity as a control parameter and can be used to investigate the interactions between auditory and linguistic processing.
- Recognition data support the hypothesis that linguistic complexity increases cognitive load. Listeners react differently on linguistic complexity. Hearing impaired listeners seem to be much more reliant on standard word order.
- The Target Detection Amplitude (TDA) enables a bias free analysis of eye movements during sentence recognition.
- Delayed eye movements in complex sentences and in noise indicate increased cognitive load.

ACKNOWLEDGMENTS

This project is funded by the Deutsche Forschungsgesellschaft. grant no. BR3668/1-2 and RU1494/2-2. Thanks go to Rebecca Carroll, Cornelia Hamann, Mirko Hanke, and Esther Ruigendijk for developing the sentence material and conducting the plausibility study.

REFERENCES

Brand, T. and Kollmeier, B. (**2002**). “Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests” J. Acoust. Soc. Am., **111**, 2801-2810.

Efron, B. and Tibshirani, R. J. (**1993**). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.

Gibson, E. (**1998**). “Linguistic complexity: locality of syntactic dependencies” Cognition, **68**, 1-76.

Kim, S. Kim, M. and Chun, M. M. (**2005**). “Concurrent working memory load can reduce distraction.” in *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, **102**, 16524-16529.

Knoeferle, P., Habets, B., Crocker, M. W., and Müntje, T. F. (**2008**). “Visual Scenes Trigger Immediate Syntactic Reanalysis: Evidence from ERPs during Situated Spoken Comprehension” Cerebral Cortex, **18**, 789-795.

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

Schriefers, H., Friederici, A. D., and Kühn, K. (**1995**). “The processing of locally ambiguous relative clauses in German” Journal of Memory and Language, **34**, 499-520.

Tanenhaus, M. K., Spivey-Knowiton, M. J., Eberharda, K. M., and Sedivy, J. C. (**1995**). „Integration of Visual and Linguistic Information in Spoken Language Comprehension” Science **268**, 1632–1634.

Tewes, U. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene - Revision 1991 (HAWIE-R)*. Bern, Stuttgart, Toronto: Huber.

Uslar, V. N., Ruigendijk, E., Hamann, C., Brand, T., and Kollmeier, B. (**2011**). “How does linguistic complexity influence intelligibility in a German audiometric sentence intelligibility test?” Int. J. Audiol., **50**, 621-631.

Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Kollmeier, B., and Brand, T. (***in preparation***). “On the interdependence of linguistic complexity, listening condition and cognitive abilities in speech intelligibility measurements: findings for listeners with normal hearing using new sentence material”

Uslar, V. N, Kollmeier, B., Brand“, T. (***in preparation***). “On the interdependence of linguistic complexity, listening condition and cognitive abilities in speech intelligibility measurements: findings for older listeners with and without hearing loss”

Wendt, D., Kollmeier, B., and Brand, T. (***in preparation***). “How does linguistic complexity influence eye movements during the recognition of sentences in an audio-visual task ?”

Wendt, D., Kollmeier, B., and Brand, T. (***in preparation***). “The influence of linguistic complexity and noise on speech comprehension: Evidence from eye movements”

Speech intelligibility in fluctuating maskers

INGA HOLUBE

Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, D-26121 Oldenburg, Germany

Within several experiments, the influence of different maskers on the speech reception threshold (SRT, signal-to-noise ratio for 50% speech intelligibility) was examined using the Oldenburg sentence test (OLSA). The maskers were stationary noises, speech or speech-like signals. The speech and speech-like signals were intelligible or non-intelligible, composed of different languages with natural or destroyed fine structure (ICRA5-like) but similar pause durations and long-term average speech spectra (LTASS). The SRT differences for normal-hearing German listeners, normal-hearing foreign native listeners and hearing-impaired Germans were small with stationary noises, but enlarged with fluctuating maskers. Intelligibility of the masker increased the SRT only slightly, whereas the ICRA5-like maskers resulted in a significant SRT increase. SRT also increased for an older normal-hearing listener group compared to a younger listener group. Composition of same or different speakers to babble noise increased the SRT even beyond its stationary noise value. Different masker levels showed a significant effect on the SRT for fluctuating maskers. Open (free oral response) and closed (response on a touch screen) test settings led to significant differences for the fluctuating masker but not for the stationary maskers. Additionally, measured reaction times for the vocal response and subjective listening effort ratings in some of the experiments were related to speech intelligibility results and independent of masker type.

INTRODUCTION

Speech intelligibility in background noise has been investigated in many studies (see Bronkhorst, 2000, for a review). Several features of the background noise, e.g., long-term average speech spectrum (LTASS), temporal gaps, fine structure, speaker sex, number of speakers, and intelligibility, influence the speech intelligibility results. Previous studies are difficult to compare because of differences in measurement methods, signals presented and subject groups. Therefore, this study used the same speech test in several experiments applying different maskers and subject groups. These experiments were an extension of the studies of Wagener and Brand (2005) and Wagener *et al.* (2006), who found a 14 dB lower speech reception threshold (SRT) for a fluctuating masker relative to a stationary masker for normal-hearing listeners, but less benefit of the temporal gaps for hearing-impaired subjects, and a higher variability in the results for fluctuating maskers. Parts of the data were published in Holube *et al.* (2009), Taesler and Holube (2009) and Holube *et al.* (2011).

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.