Influence of memory effects in speech intelligibility tasks

Stefanie Keller[1,*], Christian Wirtz[2], Hanna Beike[3], and Werner Hemmert[1]

[1] *BAI, Bio-inspired information processing, IMETUM, Technische Universität München, Germany*

[2] *MED-EL Deutschland GmbH, Germany*

[3] *Hochschule für angewandte Wissenschaften, FH München, Germany*

Testing speech reception thresholds of hearing-impaired patients is a common task in clinical routine and research. Tests consist of grammatically correct sentences containing different grammatical classes. It is expected that due to primacy and recency memory effects error rates of the first and last word are minimal. In addition, from a linguistic point of view, not only the position of a word but also its grammatical class causes different cognitive effort. This study analyses the effect of different conditions on the comprehended words belonging to different grammatical classes. So far, nine normal-hearing subjects were measured via headphones with a German speech intelligibility test with different kinds of noise and different interaural time differences. The results do not only show the expected memory effects for the noun at the first and last position of the sentences. Also significant differences for the comprehension of sentence-centered numerals were found in comparison to neighboring positions. This is impressive because in the middle, normally the attention of a listener is minimal, therefore one would expect a small recognition rate. In summary, we conclude that careful analysis of speech-reception tests also provides information on more cognitive aspects involved in speech understanding like memory capacity.

INTRODUCTION

Speech-intelligibility tasks are a common tool to measure the speech reception threshold (SRT) in noise of hearing-impaired persons. They are well-established in different Western European languages and are mostly all designed the same way: Subjects listen to a sentence in the specific language in background noise. Then they repeat all words they understood. Depending on the number of correctly understood words the signal-to-noise level is varied to determine the so-called SRT where 50% of the words were understood. For the German language, the test of choice is the *Oldenburger Satztest (Olsa)* (Wagener *et al.*, 1999).

The test consists of 40 lists composed of 30 five words sentences (Wagener *et al.*, 1999). Sentences are non-sense sentences with identical grammatical structure. As

*Corresponding author: stefanie.keller@tum.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

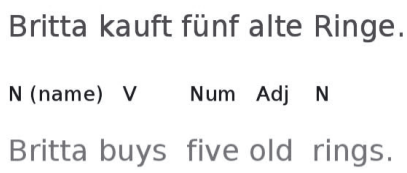

Fig. 1: Example of a typical Olsa-sentence. Upper line shows the German sentence, middle line the grammatical classes, lowest line shows the English translation.

all words can be exchanged, there is no contextual information to guess them. Each position of the sentence is filled with a word belonging to a different grammatical class (cf. Fig. 1).

Each grammatical class exhibits different cognitive effort and requires therefore different complexity for processing.[1] Verbs for example carry information about person, numerus, tempus, genus verbi, and modus, nouns about numerus, genus, and case. For this study, we will assume that nouns are more simple than verbs and adjectives because of the restricted use of objects and names. Nouns coding an object are one of the earliest words in language acquisition (Dittmann, 2002), this is a hint that these are simple words.[2] Personal names have restricted use and are sometimes triggered by personal experiences (e.g., someone might know a 'Britta' who is a very kind person). Numerals are a special group of words; it is a collection of different kinds of numbers: cardinals, ordinals, fraction numbers, etc. Here, we will only take a look at the cardinals because that is the group used in Olsa-sentences. So it is a relatively small group of words that is closed, i.e., no new speech material joins the group. That means this position in Olsa-sentences is filled by a collection which is small and predictable and should therefore be easy to classify and remember. Hence, we derive **Hypothesis 1**: The nouns (names and objects) should be recalled best, whereas a decreasing recall should be found for numerals, adjectives, and verbs (cf. Fig. 2). This should be reflected in a stable recognition rate of each grammatical class over conditions, i.e., the values shall not differ significantly from 50%, which is the value for the determined SRT.

Syntactic structure is another point that can influence speech understanding. Carroll

[1]The following reflections are mainly German specific and may just be transferred to other languages with constraints.

[2]Because of this restricted use of the grammatical class *nouns* in Olsa-sentences complex nouns like abstract words *freedom*, *wisdom*, etc. will not be considered in the following text. That is why the assumption of the smaller complexity of used nouns can be made. Note that this is not true for all representatives of the grammatical class *noun*, for further information and detailed discussion see Leiss (2002) and Vigliocco *et al.* (2011).

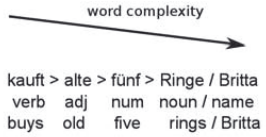

Fig. 2: Hypothesis 1: Increasing complexity for the grammatical classes used for the Olsa-sentences.

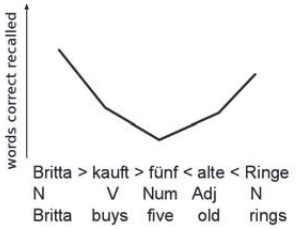

Fig. 3: Hypothesis 2: Expected primacy and recency effects for the Olsa-sentences.

and Ruigendijk (2013) found that the syntactic complexity can influence intelligibility in noise, and Uslar *et al.* (2011) showed the dependency of syntactic complex sentences on speech recognition in younger listeners. The default syntactic structure for German main clauses is subject followed by verb followed by an object, which is also known as SVO. If there is an unusual position like OVS, this causes more cognitive effort and therefore is more difficult to understand in noise (Carroll and Ruigendijk, 2013). Because the default structure is used in all Olsa-sentences, the syntactic structure will not be further analysed.

Although Carroll and Ruigendijk (2013) could not prove a connection between memory load and intelligibility, Ljung *et al.* (2013) on the other hand demonstrated that people with a higher working-memory capacity can recall words better in noise than people with a lower working-memory capacity. As postulated by Miller (1956), the working memory consists of 7 ± 2 items (Miller's law).[3] By chunking it is of course possible to remember for example a telephone number with more than nine numbers. But if you investigate the working memory, the chunks will not exceed 7 ± 2. Olsa-sentences are composed of five words, which should minimize memory effects.

Not only the grammatical class differs for cognitive load but also the position in the sentence is important. That is why we should have a look at memory effects as well. There are serial-position effects for memorising items of a list, they show a U-shaped curve: Items at the beginning and at the end tend to be better recalled (e.g., Jones and Oberauer (2013) and Oberauer *et al.* (2003)).

Investigations of these so-called primacy and recency effects have been made. Objects in the middle are not as well remembered as first and last position but the occurrance of primacy or recency effect depends on the task (Healy *et al.*, 2000). It is therefore likely that the score for the remembered words are higher at the beginning and at

[3]Working memory is a highly complex subject and cannot be fully discussed in this article.

the end of sentences. For Olsa-sentences these positions are filled by a name and an object.

According to memory load we therefore formulate **Hypothesis 2**: Recall for the first and last position of the sentences should be higher than for all other positions (Fig. 3).

METHODS

The SRT was determined with Olsa. Nine normal-hearing listeners (29.2 $\pm$ 5.2 years) were tested via earphones (Sennheiser HDA 280) in an auditory booth (IAC 350). We introduced different interaural time differences (ITDs) for speech and noise (ITDs were 0, 200, 400, and 600 μs): the speech was shifted to the side whereas the noise was presented with $0 -$ µs ITD on both channels of the earphones. The background noise was Fastl-Noise (Fastl, 1987) or olnoise, which is a noise that was produced by overlaying all Olsa-sentences, i.e., the spectrum of the noise is the spectrum of all sentences. Whereas olnoise has no additional temporal modulations, Fastl-noise is modulated similarly to speech. The conditions S0N0, S0N0$_{Fastl}$, S200N0, S400N0, S600N0 were tested (the numbers indicate the ITD in microseconds).

RESULTS

SRTs for the different conditions are shown in Fig. 4. The participants profit if the speech signal is presented from the side and the noise is presented frontally because the SRT values for the SΔtN0 are better (i.e., more negative) than in the S0N0 condition. If the noise is changed from olnoise to Fastl-Noise, subjects are able to listen into the dips; SRT values are almost 9 dB better.

Hypothesis 1

Increasing complexity should impact recognition scores in all measured conditions. Comparing the boxplots for each grammatical class and condition (Fig. 5) there are some apparent fluctuations: medians of numerals differ between conditions, sometimes the median values are above and sometimes below the 50% line, whereas these fluctuations could not be found for the other grammatical classes. An analysis of variance showed that none of these differences between the recognition rate for either grammatical class over conditions was significant, especially not for the fluctuations of the numerals ($F(4,40) = 1.62$, n.s.). The conditions for an analysis of variance are accomplished: Bartlett's test and Shapiro-Wilk's (Shapiro-Wilk's gives very robust results) test showed no significant results so that we assume equal variances across groups (homoscedasticity) and normal distribution of samples. This ANOVA result means that the recognition rate is stable over different conditions as expected.

Furthermore, adjectives and verbs are as expected the gramatical classes with the lowest medians. They show over all conditions medians lower than 50%, whereas the recognition rates of nouns (names and objects) are always (except in one case) over 50%. The following range of means of grammatical classes from low to high can

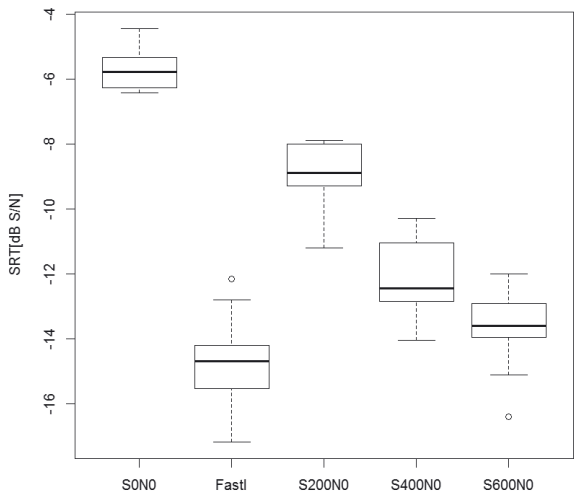

Fig. 4: Boxplots for the different spatial conditions for nine normal-hearing listeners.

be found: Adj $<$ Verb $<$ Num $<$ Noun (cf. Table 1). This is almost the predicted range for Hypothesis 1, except that verbs and adjectives changed places.

Hypothesis 2

The first and last position have the best score of recognition, cf. exemplarily Fig. 6 and 7. In the mid-position an increasing value for numerals can be seen in comparison to the neighbouring positions. Therefore a U-shaped form of recall cannot be reported because the values for the middle position are not lower or not even as low as for verbs and adjectives. The values in Table 1 do not show the predicted U-shaped form but rather a W-shaped form, where the midst value is increasing a bit.

The conditions for an ANOVA were accomplished; the analysis shows the reported behaviour. There is a significant effect of position ($F(4,20) = 36.89$, $p < 0.01$). Post hoc Tukey's test showed significant differences at $p < .05$ level between the recognition rate for the first position against all positions except the last. The rate of the middle position differs significantly from neighbouring positions. This means that the middle position can be well remembered, which was not expected according to Hypothesis 2. The second and fourth position do not show a significant difference from each other.

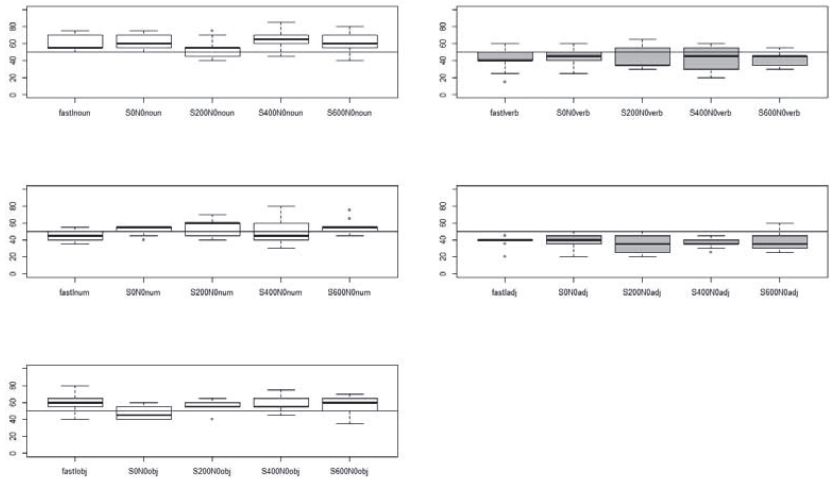

Fig. 5: Boxplots show percent correct recalled words. Each subplot shows the values for one grammatical class for the different conditions. From upper left to bottom right: nouns, verbs, adjectives, numerals, objects.

DISCUSSION

The results show an interaction between Hypotheses 1 and 2. This can be seen by the values of the middle position (numerals). Although the differences between the values for the numerals over coniditons were not significant, they show the most flucutations. Due to the grammatical class hypothesis it was not expected that numerals would show this variances over conditions in recall, and due to Hypothesis 2 it was not expected that they would show such a great recognition rate.

The results therefore show that grammatical class has an effect on the recognition rate and should be further analysed. One goal should be to clarify the explicit effect of grammatical classes by testing grammatical class and recency/primacy effects separately, e.g., by constructing lists of different grammatical classes to get rid of the syntactic structure (i.e., in this case position) or by replacing names and objects in the sentences with more difficult word classes, e.g., pronouns. Thereby the first and last position of the sentences are not filled with the least complex grammatical class.

Likewise working memory capacity should also be registered for each patient to see if his capacity is low or high. If a patient can only remember at most five things, it will be very hard to solve a task like Olsa where five items are covered with noise.

Another important fact that has to be considered is age. Larsby *et al.* (2005) and Uslar

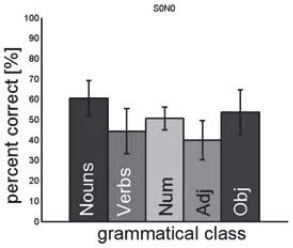

Fig. 6: Expected primacy and recency effects for the Olsa-sentences, condition S0N0.

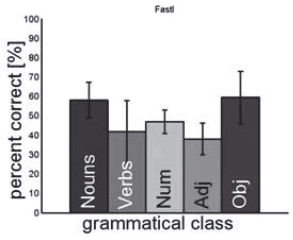

Fig. 7: Expected primacy and recency effects for the Olsa-sentences, condition S0NFastl.

Table 1: Means and standard deviations for the recognition rate for each grammatical class per condition in percent [%]. Note: SD for ALL is the standard deviation for the mean ALL.

Condition	Nouns (names)	Verbs	Num	Adj	Nouns (obj)
S0NFastl	59.44± 9.5	41.67 ± 14.79	45.55 ± 6.8	38.33 ± 7.5	60 ± 12.75
S0N0	62.22 ± 9.39	45 ±10.6	51.11 ± 5.46	39.44 ±9.17	48.3± 8.3
S200N0	54.44± 11.57	42.78 ± 13.5	56.11 ± 10.6	35 ± 10.6	56.11 ± 6.97
S400N0	65 ± 10.6	43.33 ± 6.0	50.55 ± 14.36	36.11 ± 15.7	57.77 ± 9.05
S600N0	60.55 ± 12.6	42.77± 7.95	55.55±9.17	38.89 ± 11.11	56.11 ± 11.93
ALL	**59.13**± 3.69	**43.50** ± 1.5	**52.77** ± 4.2	**37.55** ± 1.9	**55.38** ±4.11

et al. (2011) found that elderly people have more difficulties than younger people to understand speech in noise. As typically hearing-impaired people are older, they may not perform as good as teenage/young-adult groups of hearing-impaired people.

For a detailed analysis or conclusion of this interaction, it is required to test more subjects of different age, including their memory capacity.

We have shown that with a careful analysis we can evaluate working-memory effects already with a standard SRT test.

ACKNOWLEDGEMENTS

This work was supported by the Munich Bernstein Centre for Computational Neuroscience (01GQ1004B and 01GQ1004D) and MED-EL.

REFERENCES

Carroll, R., and Ruigendijk, E. (**2013**). "The effects of syntactic complexity on processing sentences in noise," J. Psycholinguist. Res., **42**, 139-159.

Dittmann, J. (**2002**). *Der Spracherwerb des Kindes. Verlauf und Störungen.* (Beck Verlag, München), 1. Auflage.

Fastl, H. (**1987**). "Ein Störgeräusch für die Sprachaudiometrie," Audiol. Akustik, **26**, 2-13.

Healy, A.F., Havas, D.A., and Parker, J.T. (**2000**). "Comparing serial position effects in semantic and episodic memory using reconstruction of order tasks," J. Memo. Lang., **42**, 147-167.

Jones, T., and Oberauer, K. (**2013**). "Serial-position effects for items and relations in short-term memory," Memory, **21**, 347-365.

Larsby, B., Hällgren, M., Lyxell, B., and Arlinger, S. (**2005**). "Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects," Int. J. Audiol., **44**, 131-143.

Leiss, E. (**2002**) "Die Wortart Verb", in *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen. XVII. Die Architektur des Wortschatzes I: Die Wortarten. 2. Halbband*. Edited by D.A. Cruse, (de Gruyter, Berlin, New York), pp. 605-616.

Ljung, R., Israelsson, K., and Hygge, S. (**2013**). "Speech intelligibility and recall of spoken material heard at different signal-to-noise ratios and the role played by working memory capacity." Appl. Cognitive Psych., **27**, 198-203.

Miller, G.A. (**1956**). "The magical number seven, plus or minus two. Some limits on our capacity for processing information," Psychol. Rev., **63**, 82-97.

Oberauer, K. (**2003**). "Understanding serial position curves in short-term recognition and recall," J. Mem. Lang., **49**, 469-483.

Uslar, V., 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.

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H., and Cappa, S.F. (**2011**). "Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies," Neurosci. Biobehav. R., **35**, 407-426.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test," Z. Audiol., **38**, 5-15.