



Does sentence complexity interfere with intelligibility in noise? Evaluation of the *Oldenburg Linguistically and Audiologically Controlled Sentence Test (OLACS)*

Verena N. UsLAR¹, Thomas Brand¹, Mirko Hanke², Rebecca Carroll², Esther Ruigendijk²,
Cornelia Hamann², Birger Kollmeier¹

¹Institute of Physics, CvO University of Oldenburg, Germany

²Department of Modern Languages, CvO University of Oldenburg, Germany

verena.uslar@uni-oldenburg.de

Abstract

The *Oldenburg Linguistically and Audiologically Controlled Sentence Test (OLACS)*, which contains sentences with seven different grades of linguistic complexity, is introduced. The evaluation of this new German speech material was performed by presenting each sentence at three different SNRs to 36 normally hearing listeners. Sentence specific discrimination functions were calculated and for each sentence type 40 sentences were selected. Differences of up to 3dB occurred for the different grades of linguistic complexity. Interindividual differences occurred in speech recognition rates of up to 30%. Thus, on the one hand the material seems to be appropriate for examining the influence of sentence complexity on speech recognition both qualitatively as well as quantitatively. On the other hand the OLACS might be used for diagnostic purposes to differentiate e.g. across individual listeners.

Index Terms: speech recognition, linguistic complexity, noise

1. Introduction

Understanding speech in acoustically difficult settings (“Cocktail party effect” [1]) is one of the most important abilities used in human communication. So far, the underlying mechanisms have not been understood sufficiently. By joining audiological and psycholinguistic expertise, the AULIN project aims at obtaining both a qualitatively and quantitatively better understanding of the processes involved in speech processing in noise.

It is known that, besides sensory factors, cognitive factors related to, for instance, age, working memory or attention seem to play a role in understanding speech, especially in fluctuating noise [2]. Also, linguistic complexity (in particular structural, or syntactic complexity) is an influential factor in speech processing and it seems to interact with speech recognition [3], possibly because understanding more complex speech relies more heavily on working memory capacity. Furthermore, linguistic complexity interacts with other factors, such as age or hearing impairment [4]. So far, the existing material used in German speech intelligibility tests does not have the controlled and carefully graded linguistic complexity needed for studying the effect of linguistic complexity on speech recognition. Therefore, we developed a German speech intelligibility test containing test lists with graded linguistic complexity, the so called *Oldenburg Linguistically and Audiologically Controlled Sentence Test (OLACS)*.

This article describes the characteristics of the new sentence material and presents the results of its evaluation. Since

differences between the sentence types of different linguistic complexity were to be expected and wanted, one problem was to concurrently assure that all differences in intelligibility across the different sentence types do indeed originate from the different sentence structures and not from differences in sensory/acoustic factors. The primary goal was assessing the recognition rates for each sentence in order to establish homogeneous test lists, but we hypothesised that the following two effects should be observable during our measurements:

1. There should be differences between sentence types regarding their intelligibility.
2. There should be differences between listeners based on their individual ability to understand the different sentence types in noise.

2. Sentence material

The sentence material for OLACS consists of seven types of sentences with graded linguistic complexity. The syntactical structure is identical within each condition. The evaluation was started using 720 sentences. Overall 114 different nouns (i.e. job titles like “baker”, animals like “elephant”, and fantasy figures like “faeries”), 45 simple adjectives (i.e. large, green, old, etc.) and 36 verbs were used. All words were selected regarding their presentability in drawings (important for planned measurements with an eye-tracker also done in this research project, [5]) and their frequency in the German language. Using this fixed set of words ensured a relatively equal phonem distribution over all sentence types.

Each sentence falls into one of seven different types. Three of these sentence types are of the “Verb2” category (verb in the second position of the sentence, see Tab. 1, the three uppermost examples) and four of these types use embedded relative clauses (Tab. 1, the four lower examples). The Verb2 type with the subject-verb-object structure (SVO) is the most usual sentence structure for German sentences [6]. As German possesses overt (and in the case of male nouns mostly unambiguous) marking of grammatical case on the articles, it is possible to build sentences with an object-initial word order (Object-Verb-Subject, OVS) without using passive verb forms, which still convey an unambiguous meaning of the sentence. In German everyday speech the OVS sentence structure is far less common than the SVO structure. Therefore, the OVS structure is expected to be harder to understand, because potential listeners are not as familiar with it as with the SVO structure. The German article “die”, which denominates the nominative as well

3. Methods

Table 1: Examples for the seven types of sentences. SMALL CAPS-TALS indicate the point of disambiguation in each sentence type. *Nom* (nominative), *acc* (accusative) and *amb* (ambiguous case) indicate the case of the words. *Pl* indicates plural forms.

SVO	DER	kleine	Junge	grüsst	den	lieben	Vater.
	THE	little	boy	greet	the	nice	father
	<i>nom</i>	<i>nom</i>			<i>acc</i>	<i>acc</i>	
OVS	DEN	lieben	Vater	grüsst	der	kleine	Junge.
	THE	nice	father	greet	the	little	boy
	<i>acc</i>	<i>acc</i>			<i>nom</i>	<i>nom</i>	
amb	Die	liebe	Königin	grüsst	DER	kleine	Junge.
OVS	the	nice	queen	greet	THE	little	boy
	<i>amb</i>	<i>amb</i>			<i>nom</i>	<i>nom</i>	
SR	Der	Bauer,	DER	die	Lehrer	fängt,	lacht.
	the	farmer	WHO	the	teachers	catches	smiles
			<i>nom</i>	<i>amb</i>			
OR	Der	Bauer,	DEN	die	Lehrer	fangen,	lacht.
	the	farmer	WHO	the	teachers	catch	smiles
			<i>acc</i>	<i>amb</i>		<i>pl</i>	
amb	Die	Bauern,	die	die	Lehrerin	FANGEN	lachen.
SR	the	farmers	who	the	teacher	CATCH	smile
			<i>amb</i>	<i>amb</i>		<i>pl</i>	<i>pl</i>
amb	Die	Bauern,	die	die	Lehrerin	FÄNGT	lachen.
OR	the	farmers	who	the	teacher	CATCHES	smile
			<i>amb</i>	<i>amb</i>			<i>pl</i>

as the accusative of female nouns, is making it possible to create OVS sentences with late disambiguation. Therefore, in the case of the ambiguous object initial sentences (amb OVS) it gets clear just with the utterance of the article “der” that the part before the verb (i.e. “Die liebe Königin”, engl. “the nice queen”) is indeed the object of the sentence. Since a listener may reckon that the usual SVO structure applies here the utterance of the “der”, which clearly indicates the nominative, makes it necessary to revise any initial hypotheses about sentence structure and meaning. This necessary restructuring may lead to an increase in processing cost.

In the German language, relative clauses may either be subject relative clauses (SR, the more common structure) or object relative clauses (OR). Both sentence structures can be build with and without ambiguity in German. Consequently, the second category of sentences using relative clauses contains four different sentence types. In general, SR sentences should be easier to process than OR sentences, because grammatical function (and/or thematical role, animacy, etc.) relations remain constant between matrix clause and embedded clause in the former, but not in the latter type of sentence. The relative pronoun in German has the same shape as a definite article and inflects for case and number of the noun phrase it refers to. For masculin singular nouns this inflectional marking is unambiguous and allows for an immediate correct interpretation of the grammatical role relations within the sentence. In the ambiguous relative clause conditions the relative pronouns have the same form for both nominative and accusative case (“die”). It is then only at the verb of the relative clause, that the correct grammatical function and semantic role configuration can be determined. As with our OVS sentences, this late disambiguation should make it necessary to revise initial structural decisions and lead to greater processing cost. These assumed systematic differences in the processing effort for the different sentence types were used as the independent variable in the experiments described here.

3.1. Test Design

The evaluation of the sentence specific intelligibility for the whole material took place in three consecutive measurement phases. During evaluation phase I, the recognition rates were measured at -7 dB SNR for fragments of 720 sentences which resulted from cutting each sentence into three pieces. Verb2 sentences were cut directly before and after the verb yielding for example fragments like: “Der nette Elefant” (“the nice elephant”), “umarmt” (“hugs”), and “den kleinen Igel” (“the small hedgehog”). Sentences with embedded relative clauses were cut at both commas, yielding fragments like “Der Zauberer” (“the wizard”), “der die Zwerge interviewt” (“who the dwarfs interviews”) and “lacht” (“laughs”).

This was done to first measure the intelligibility based on acoustical cues while excluding syntactical and context effects. For example, the word “Elefant” (elephant) was part both of a SVO sentence (i.e., placed at the beginning of the sentence) and part of an OVS sentence (i.e., at the end of a sentence) as well as a part of sentences with a relative clause (either placed in the main sentence or in the relative clause), respectively. If the “Elefant” is presented within a whole sentence of any type it remains unclear if a possible difference of the intelligibility of the word in the respective type of sentence is mainly caused by the respective acoustical representation or the syntactical structure. By presenting the sentence fragments in evaluation phase I we eliminated the effect of the sentence structure and therefore focused on acoustical cues. After evaluation phase I all sentences where there was one or more words which were never understood correctly by any of the listeners were discarded to make the material more homogeneous and to keep the total measurement time as short as possible.

In evaluation phase II the intelligibility of the selected 560 sentences was measured, again at -7 dB SPL, this time for the whole sentences. After evaluation phase II the mean intelligibility and the standard deviation for each type of sentence was calculated. Sentences, that deviated in intelligibility by more than two standard deviations for the respective type of sentence, were discarded. Also the results of evaluation phase I were used to compare the recognition rates for each fragment with the mean recognition rates for the respective fragment type and with the recognition rate when the whole sentence was presented. In evaluation phase III the recognition rates at two sentence specific SNRs were determined for the remaining 360 sentences with best match in intelligibility and sentence specific discrimination functions were calculated.

3.2. Listeners

36 normally hearing listeners participated voluntarily in the evaluation of the OLACS test material, with 12 different listeners in each phase of the evaluation. The listeners were between 20 and 30 years old (mean and SD: 25 ± 3.5) and they were all students at the University of Oldenburg. Each listener received 10 Euros per hour for the participation in this study.

3.3. Test Setup

Sentences or sentence fragments were presented in a random order in a soundproofed chamber via Sennheiser HDA 200 headphones. The noise was presented at a sound pressure level of 65 dB SPL and the speech was presented according to the parameters described in 3.1. After presentation of each sentence, the participant was asked to repeat what she/he had just heard.

Table 2: mean recognition rates [%] in evaluation phase I (i.e. fragments, upper panel) and evaluation phase II (i.e. whole sentences, lower panel) averaged across listeners for each word (i.e. W1 = first word in each sentence; W7 = seventh word) and for the whole condition (cond).

	W1	W2	W3	W4	W5	W6	W7	cond
SVO	81	69	62	26	44	36	47	57
OVS	40	49	56	32	84	64	58	55
amb OVS	77	64	57	34	89	69	67	66
SR	78	71	77	66	49	44	72	65
OR	76	70	44	63	48	25	79	58
amb SR	60	44	23	83	67	22	72	53
amb OR	62	48	18	81	67	48	73	57
SVO	84	73	70	60	61	29	31	58
OVS	62	58	68	52	48	26	24	48
amb OVS	82	72	72	56	40	23	32	54
SR	90	86	80	80	60	49	70	74
OR	87	86	52	79	59	34	67	66
amb SR	81	65	73	75	63	32	53	63
amb OR	83	64	73	77	65	42	53	65

Participants were explicitly allowed to guess. The investigator, who was located in the booth together with the participant, marked the correctly repeated words on a computer screen. All seven words were included with equal weighting in the calculation of the intelligibility. A word was counted as correct only if the correct form (that is, with correct case, gender, number, and tense) was used. The first 20 presentations were regarded as training trials and the results were discarded. A pause of at least 5 min was made every 20 to 30 minutes. Measurements were resumed on another day, if they exceeded two hours.

4. Results

4.1. Analysis of the differences between sentence types

The upper panel in Table 2 shows the recognition rates grouped by sentence type for all listeners for each word (W1 to W7) and over the whole condition in evaluation I (i.e. sentence fragments). For instance, the first article (W1) in the SVO condition was understood correctly in 81% of all presentations of a subject fragment of a SVO sentence. The recognition rates for the subject fragments and the ambiguous object fragment (which could have been a subject to) are comparable in all three conditions (light gray cells in the upper panel). Non-ambiguous object fragments are significantly less frequently recognised (compare light gray and medium gray cells of the upper panel). The ambiguous OVS sentences are in average better recognized than the SVO and OVS sentences. The comparisons between the second and fifth word in the relative clause conditions show a clear preference for the singular/male form of a noun. In the ambiguous sentences the second word is always a female noun or a male noun in plural form, whereas in the fifth word it is the other way round. The singular/male form gets about 20% higher recognition rates (e.g. the recognition rates of the dark gray cells in upper panel). This preference for the singular form also occurs in the sixth word (relative clause verb). In the OR and the ambiguous SR condition, where the verb is in its plural form, the recognition rate is about 20% lower.

The lower panel in Table 2 shows the results for evaluation phase II (i.e. whole sentences). The first words show the highest recognition rates in each condition and for each following word the recognition rate decreases, with an exception for the

verb of the main sentence (primacy effect and recency effects, i.e. the well documented effect, that the early words or digits in a row are easier to remember than the middle ones and that the last items in a row are again easier to remember [7]). The comparison between the different relative clauses shows, that there is an effect of ambiguity at least for both SR types (mean recognition rate of amb SR is 11% lower compared to simple SR). For the OR sentences no such effect is visible. The preference for male/singular forms is again apparent (e.g. dark gray cells in lower panel). Sentences of the Verb2 type have an average recognition rate of 53% and sentences with embedded relative clauses have an average recognition rate of 67%, with the SR structure yielding the best recognition rate (74%). The recognition increases slightly if the whole sentence is presented. The OVS and the ambiguous OVS sentences are the exception to this. Here the recognition of the subject part of the sentences (W5 to W7) decreases to about half the value of the recognition rate when the fragment is presented alone (light gray cells of upper and lower panel).

The 50% speech reception threshold (SRT) and the slope of the discrimination function of each sentence was approximated with a maximum likelihood algorithm using three recognition rates measured at three different SNRs. Since there still was a large range for the calculated SRTs and the slopes of the functions we again discarded about 10 sentences per condition. This time we took into account the deviation of the SRT of the sentences from the mean of the respective type of sentences and the overall phonem distribution in each condition, which should match the phonem distribution for the German language as closely as possible. In the end we had at least 40 sentences per condition. By discarding roughly 3/4 of all sentences the number of different nouns in the material was reduced from 114 to 105. The number of verbs and adjectives stayed the same.

Figure 1 shows the mean discrimination functions for the seven sentence types. They were determined by using all recognition rates measured at the respective SNRs during the second and third evaluation phase for the remaining 40 sentences in each respective condition to approximate the respective mean discrimination function. The SR condition shows the lowest SRT and the OVS condition yields the highest SRT (difference in SRT: 2.6 dB). Also, the ambiguous SR sentences show a higher SRT than the respective non-ambiguous type. The ambiguous OVS sentences show the same SRT as the SVO sentences but a significantly shallower slope than all other sentence types. All this is in accordance with the results presented in Table 2.

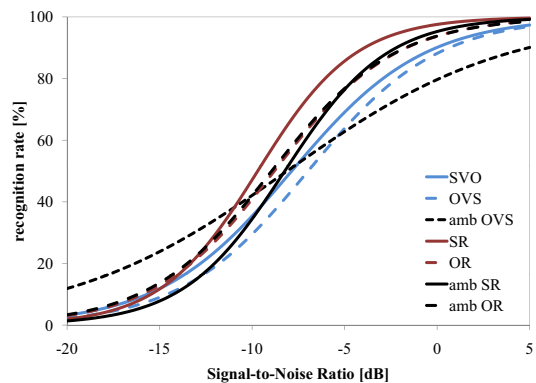


Figure 1: mean sentence discrimination functions (recognition rate [%] over signal-to-noise ratio [dB]) for each sentence type.

4.2. Analysis of the differences across listeners

Figure 2 shows the recognition rate for each condition for five representative listeners in evaluation phase II. For the purpose of clarity and since all listeners showed the same characteristics only the results of five out of 12 listeners are depicted here. Listener BL88 (blue bars) recognized the most of all with over 70% correct responses over all conditions whereas listeners PL85 (red bars) only recognized about 46% of all words in all conditions correctly. In general, listeners performed either good or bad in any condition. However, the variability across listeners depends on the condition. The smallest differences between listeners occur in the SVO structure (about 15% between the best and the worst listener) whereas the OVS structure yielded the biggest differences (about 35% across listeners).

5. Discussion

The differences in the word recognition rates between evaluation I and evaluation II show the importance of eliminating the effect of the sentence structure in the first step of the evaluation. It was only possible to distinguish between the effect of the respective acoustical representation and the effect of the syntactical structure by comparing the results of evaluation I and II, which in turn enabled us to reliably discard sentences based on their acoustical/sensorial divergence. Also, it gave more insight into the effect of the changing sentence structure.

In contrast to our first assumptions, sentences with embedded relative clauses were better recognized than Verb2 sentences. This may be explained by the higher information density and the additional memory load introduced by the adjectives in the Verb2 conditions. Additionally, the slight pauses in the embedded relative clause conditions provided by the commas, may be used by the listeners to recollect the heard words and to better memorise the sentence. Contrary to the other conditions, OVS and ambiguous OVS structures get higher recognition rates when presented in fragments rather than whole sentences. When presented in fragments, the subject fragments are understood well, but when presented in whole sentences, clearly the effect of the unusual sentence order, the ambiguity and the primacy effect are taking over since in both conditions the “preferred” subject stands at the end of the sentence. The primacy effect also explains why there is no difference in the recognition rate between SVO and ambiguous OVS sentences. At the point where the primacy effect takes hold (namely at the fourth or fifth word), the listeners still perceive the amb OVS structure as a SVO sentence.

The effect of ambiguity is visible by comparing the SR and amb SR conditions, both in their overall intelligibility as well as the mean SRT, respectively. When comparing the OR and the amb OR conditions and the OVS and amb OVS conditions, respectively, the effect of the ambiguity is only visible in the higher mean SRT of the amb OR condition and the shallower slope of the amb OVS condition. In all conditions the ambiguity increases the differences in intelligibility between the listeners, indicating that for more capable listeners this additional difficulty poses less of a problem.

For practical purposes, for each of the seven conditions, two test lists with 20 sentences each which are equal with regards to their intelligibility and phonem distribution, have to be established. Those lists have to be tested with different groups of listeners (younger and older listeners, hearing impaired persons, children, etc.) and varying noise conditions, which might also further our understanding of speech processing in noise.

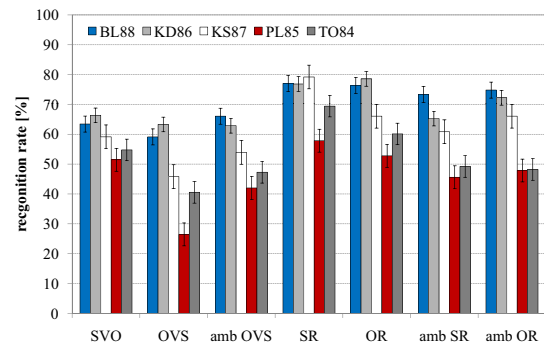


Figure 2: mean recognition rate [%] for each condition for five out of 12 listeners in evaluation phase II.

6. Conclusions

- For each of the seven sentence types a list of 40 sentences is established. Each list is highly homogeneous with respect to the intelligibility across sentence fragments and whole sentences. Moreover, they approximate the German phonem distribution.
- The complexity of the sentence structure has a significant impact on speech intelligibility even in a homogeneous group of young normally hearing listeners. This indicates that the OLACS material might be well suited for investigating the qualitative and quantitative effects of linguistic complexity on speech intelligibility in different groups of listeners.
- The large differences across listeners indicate that individual cognitive capability may play an important role in the processing of speech in noise. This effect in combination with the consistency of the rank order across conditions allows the test to be used for diagnostic purposes, e.g. to differentiate between individual listeners.

7. Acknowledgements

Supported by the Deutsche Forschungsgesellschaft (project AULIN, KO 942/20-1 and HA 2335/2-1).

8. References

- [1] Cherry, E.C., Some experiments on the recognition of speech, with one and with two ears. *JASA*, 25(5): 975-979, 1953.
- [2] George, E.L.J., Festen, J.M., & Houtgast, T., Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *JASA*, 120(4): 2295-2311, 2006.
- [3] Weik, V.N. et al., Einfluss von Satzkomplexität auf das Sprachverstehen am Beispiel des Göttinger Satztests. 12. annual meeting of the german association for audiology (DGA), Innsbruck, 2009.
- [4] Wingfield, A. et al., Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *JAAA*, 17: 487-497, 2006.
- [5] Wendt, D., Brand, T., & Kollmeier, B., Decision moments and eye movements: When is a complex sentence understood? In *Proc. Interspeech (2010)*, this issue.
- [6] Gorrell, P., The subject-before-object preference in German clauses. In: Hemforth B, Konieczny L, (eds.). *German Sentence Processing* (pp. 25-63), Dordrecht: Kluwer, 2000.
- [7] Murdock Jr., B. B., The serial position effect of free recall. *J Exp Psy*, 64(5): 482-488, 1962.