When you can't hear your own word: Producing language in noise

UNDER REVIEW – PLEASE DO NOT CITE

Mirko Hanke*, Esther Ruigendijk, Cornelia Hamann
(School of Linguistics and Cultural Studies, University of Oldenburg, Oldenburg, Germany)

* Corresponding author: *Email* mirko.hanke@uni-oldenburg.de, *Phone* +49 441 3090904

ABSTRACT

The paper reports a language production study which assesses the effects of speech-free, fluctuating *babble* noise on the speaking process. Our study focused on the effect of noise on the structural level of sentence generation, or *formulation* in terms of the language production model by Levelt (1989). More particularly, we were interested whether adverse, noisy background conditions would result in reduced structural complexity of spoken language. For the study, 12 native speakers of German were interviewed and asked to tell a picture story in both silence and under noise. The recorded speech was transcribed and manually searched for indicators of structural complexity. In contrast to previous observations by Kemper, Herman and Lian (2003), we found no indication of reduced grammatical complexity for language produced under noise. However, subjects made significantly more errors which lead to ungrammaticality. We discuss the result in terms of different potential mechanisms of interference between noise and the speaking process and address the potential role of the monitor in the on-line generation of intra-sentential dependencies.

INTRODUCTION

Speaking appears to most adults as one of the most natural things to do. We are normally able to speak in different environments, in the hushed atmosphere of a library as well as in the middle of a city ridden with traffic noise. Looking more closely, however, psycholinguists have established that speaking requires the correct function of fine-tuned processes (e. g. Kolk, 1995). The intuitive experience that talking in noise seems more cumbersome or effortful to many speakers indicates that the machinery required for speaking can get out of tune at times. Thus there may be limits to the extent that speakers can cope with distracting (environmental) stimuli. The current study set out to investigate the effects of difficult, noisy conditions on structural aspects of spoken utterances.

THEORETICAL BACKGROUND

Difficulties with language comprehension in noise appear at first sight rather straightforwardly a result of masking of the signal, as in this case the actual acoustic representation of a sentence or word is changed before it reaches the ear of the comprehender. However, processing-for- comprehension under adverse conditions with signal degradation appears particularly difficult when complex, non-canonical, or otherwise 'unusual' sentence structure is being processed. It has been argued that the additional effort necessary to restore a useful speech signal from the masked signal draws on cognitive resources otherwise needed for the comprehension of structurally complex sentences, for instance by Wingfield and Tun (1999); Wingfield, Peelle and Grossman (2003), who frame their results in terms of Patrick Rabbitt's effortfulness hypothesis (Rabbitt, 1968, 1991; also see Uslar, Ruigendijk, Hamann, Brand & Kollmeier, 2011; Carroll & Ruigendijk, in press for more recent results and similar theoretical accounts).

Speaking in noise

Previous observations. As for the influence of noise on speaking, it is certainly less straightforward to postulate a comparable mechanism. Still, noise has been shown to exert an influence on the speaking process at different levels. For almost exactly a century, one strand of previous empirical work has been concerned with acoustic or prosodic effects of noise on speaking, in studies about the so-called Lombard reflex or Lombard effect. This effect is among other things characterised by spectral changes, an increase in voice intensity and an increase in articulation effort perceived by speakers when producing language in noise (Lombard, 1911; see Junqua, Fincke & Field, 1999, or Lu & Cooke, 2008 for more recent work on this effect). These changes have been argued to be triggered by a subconsciously operating self-regulation mechanism (Lu & Cooke, 2008). However, it remains an open question, whether this kind of adaptation only pertains to the phonetic/phonological level, or if other aspects of language production might also be influenced by noise.

Previous studies that looked at the effect of noise on speaking at a sentential level pointed to another noticeable effect on speakers. In an early demonstration of the effect of noise on speaking, Hanley and Steer (1949) showed a significant reduction of speech rate in words per minute (also see Hanke, Hamann & Ruigendijk, in press, for a similar observation). Since subjects in those studies had not been trained or instructed in any particular direction, this behaviour might reflect an automatic adaptation to the environmental setting for this particular task.

Postma and Noordanus (1996) used noise to interfere with processing resources during monitoring for *speech errors*. Their results show an increase in the number of undetected phonological errors and dysfluencies under noise. The authors account for their results based on the ‘levels of integration’ language production architecture (Garrett, 1980; Levelt, 1989), more specifically on the assumption of a monitoring system that checks language production on-the-fly for errors.

The monitoring system compares different levels of structure down the line to the intended conceptual message during the production process. In case of mismatch, formulation and/or articulation can be interrupted and a repair can be attempted. According to Levelt (1983) and Levelt (1989), the monitoring component ensures error-free production with respect to the multiple levels of an utterance (Levelt, 1989: 460), and thus serves a "corrective" function (Postma, 2000). Levelt's *perceptual loop theory* (Levelt, 1983, 1989) assumes two different channels through which (parts of) an utterance can be fed back into the comprehension system. An internal loop is created by allowing a direct connection between production and comprehension tapping the phonological plan or internal speech that is output from the formulation stage. The external loop draws on overt speech that is processed through the auditory perceptual system first and then fed into the comprehension system.[1]

The external loop seems to form a prima facie plausible candidate for the necessary interface between perception and production at which external noise might exert an influence. Postma and Noordanus (1996) support this assumption with experimental data on the negative influence of noise on the external loop. According to their results, impairment of the monitor's proper function via the perceptual loop leads to higher error rates on the phonological and lexical level.

Sentence production in noise. It seems relatively uncontroversial to us to explain acoustic/prosodic and error rate effects of noise by assuming a monitoring system which operates through auditory feedback. However, results by Kemper et al. (2003) indicated that difficult acoustic surroundings might even have an effect on earlier stages of utterance generation. The authors analysed several characteristics of spoken language production by different populations under *dual-task conditions*, where the cognitive load from a secondary task limits the availability of cognitive resources for a primary task (cf. e. g. Pashler, 1994). Crucially, their study included an experimental setting where subjects were presented with ‘cafeteria babble’ background noise. Kemper et al. (2003) reported differential effects of dual-task conditions on grammatical complexity, contingent on the age group speakers belonged to. While older participants in their study relied on a reduction of the rate of speech, younger participants produced speech which was *reduced in grammatical complexity*. As measures of complexity the study by Kemper et al. (2003) used the mean number of clauses per utterance (MCU), and D-Level, a compound score of grammatical complexity based on the age of acquisition for particular types of sentence structure (Rosenberg & Abbeduto, 1987).

The authors interpret their results as evidence for the assumption that simply listening to noise can exert a dual-task effect, which can influence sentence formulation processes. A central, mediating role in the explanation of dual-task effects on speaking is played by the notion of processing or *working memory capacity*. Authors like Gibson (1998), Kemper and Kemtes (1999) and Kemper and Sumner (2001) have argued that individual capacity constraints form an important boundary condition for the processing of complex sentence structure. Kemper and Sumner (2001) investigated effects of ageing on speakers’ verbal ability, including measures for grammatical complexity. They found significant correlations with different measures for working memory capacity constructs that had been argued to be relevant in language processing. Based on their results, the authors claim that age-related reductions in cognitive capacity serves as a cause for reduced grammatical complexity found in speech produced by older speakers. Similar to the effects of ageing, Kemper et al. (2003) claim that difficult, noisy situations are taxing on processing or memory capacity and that speakers have to adjust the level of grammatical complexity during speaking as a coping strategy.

Irrelevant sound effect. An obvious question in this respect is how simply listening to or ignoring speech or noise might lead to a reduction on available working memory capacity or to an additional cognitive load comparable to other kinds of dual-task effects. In order to substantiate this claim, Kemper and colleagues point to a number of interesting observations from attention and memory psychology. Empirical research about the so-called *irrelevant speech effect* has converged upon the finding that concurrent but unattended speech has a detrimental effect on processing for tasks that involve the recalling of (unconnected) verbal material (cf. for instance Banbury & Berry, 1998; Banbury, Macken, Tremblay & Jones, 2001). Somewhat counter-intuitively, the effect occurs no matter what intensity or content the concurrent speech has, albeit the strength of the effect might differ (Klatte, Kilcher & Hellbrück, 1995; Ellermeier & Hellbrück, 1998; Salamé & Baddeley, 1982). Importantly, the effects of irrelevant auditory stimuli on verbal recall are not limited to irrelevant speech. As for instance Jones and Macken (1993) and Klatte and Hellbrück (1993) have shown, series of intermittent tones or speech-free noise lead to performance decreases. What is more, the negative effect on performance is substantially stronger with fluctuating, ‘babble-like’ noise than with constant noise (Klatte et al., 1995). Hence, these results indicate that simply listening to, or ignoring certain kinds of noise can draw on cognitive resources necessary for other tasks. Based on the interpretation of the ISE proposed for instance by Jones (1993), it can be argued that this competition for resources is similar to experiments with a dual-task paradigm.

More particularly, with their *changing state hypothesis*, Jones and Macken (1993) and Macken, Tremblay, Alford and Jones (1999) emphasise the role of the structured nature of both speech as well as fluctuating, non-speech distractor sounds. The authors suggest that any structured sound signal is automatically subjected to auditory scene analysis, which tries to extract an ordered sequence of auditory objects from the signal. The hypothesis predicts that processing for the primary task of maintaining an ordered list of words in working memory will compete for processing resources on a *seriation process*, that would be engaged concurrently by automatic auditory scene analysis.

While it is difficult to compare the typical ISE task of recalling ordered lists of words to ‘normal’ speaking, it might be the case that keeping track of the serial order of elements in one's own speech is a very basic, underlying function sub-serving the language production process as well, including monitoring of previous speech. Still, the question remains which steps or functions performed during natural language processing-for-production will be susceptible to similar effects of a dual-task load and what the mechanism for the interference could look like. As the results by Kemper et al. (2003) indicated, speakers might attempt to overcome situational difficulties by resorting to 'easier' sentence structures. This observation then poses the question how the complexity or difficulty of a particular structure and cognitive load are related in the formulation process.

Sentence structure and linguistic complexity

Syntactic structure has been at the core of research about constraints on language processing by speakers and listeners (Jackendoff, 2002). Observational studies about language acquisition and language breakdown, as well as corpus studies have indicated that certain sentence structures are used with considerably lower frequency of occurrence than others, or hardly occur at all. Formalised notions of structural complexity based on linguistic theory are sometimes used as a means to predict processing difficulty or even processing time characteristics, as exemplified for instance by Frazier (1987), Gibson (1998, 2000), and in other authors' metrics that are defined structurally. Lewis (1996) provides a survey and discussion for a number of attempts in this direction, also see Roark, Mitchell and Hollingshead (2007). One of the perhaps simplest construals of complexity is certainly based on the amount of elements in a given domain of analysis, for instance word length in phonemes or morphemes, sentence length in words or text length measured in sentences, or in other combinations of elements within a given domain of inquiry. On the level of sentences that is in the focus of the current study, this notion is the basic rationale behind the *mean length of utterance* (MLU, Brown, 1973) measure that is used to approximate grammatical complexity in young children, or the mean clauses per utterance (MCU) measure.

However, such measures might form rather coarse measuring tools when it comes to more intricate differences between sentence structures which have apparent effects on how well sentences can be processed or how difficult, acceptable, or grammatical a sentence is judged intuitively. In psycholinguistic research on special populations, structural measures based on linguistically defined metrics are relatively common. Research on typical language acquisition for instance has aimed to establish developmental pathways from 'simpler' to more 'complex' sentence structure (cf. e. g. Brown, 1973; Bloom, Lahey, Hood, Lifter & Fiess, 1980; Rosenberg & Abbeduto, 1987). These pathways can be contrasted with acquisition patterns found in children with language impairments. In this vein, research on hearing impairment and language pathologies has yielded evidence that certain types of syntactic structure are more error-prone in production, in addition to being acquired later (Brannon & Murry, 1966; Svirsky, Robbins, Kirk, Pisoni & Miyamoto, 2000; Friedmann & Szterman, 2006; Hamann, Tuller, Monjauze, Delage & Henry, 2007; Delage, Monjauze, Hamann & Tuller, 2008; Friedmann, Belletti & Rizzi, 2009; Delage & Tuller, 2010, Jakubowicz, 2010).

A critical assumption implicit in this and other kinds of psycholinguistic work is that a structure which is more 'complex' at the level of formal linguistic description will also be cognitively more costly to process (see e. g. Fanselow, Kliegl & Schlesewsky, 1999). Since complexity of sentences is not simply a function of the amount of words, other factors have to be taken into account as well, as explained in the following section.[2]

The current study

The current study aims to explore the effects of noise on linguistic complexity of spoken language in more detail, and in order to do so we used a number of different structure types as indicators for sentential complexity. In particular, we are concerned with the question whether a noise-induced secondary task load affects cognitive processes implied by language processing-for-production on the (on-line) processing of sentence structure during grammatical encoding.

With respect to the compound scores used in previous studies by Kemper and colleagues, it might be of additional interest, especially from a linguistic viewpoint, whether certain structures will be affected more than others by processing detriments related to background noise, or which aspects of sentence complexity are particularly difficult to process under difficult acoustic conditions. At the same time, even to this day it poses a major problem for linguistic and psycholinguistic theory to establish generally accepted, 'unified', and empirically solid notions of complexity. For practical reasons, in the current study we will search for different indicators of relative sentence complexity where we see converging evidence from both psycholinguistic work and from theoretical assumptions about structural complexity based on formal linguistic analysis. The structures we used have been shown to be indicative of some kind of processing difficulty in studies on language acquisition, language breakdown, as well as in on-line comprehension studies with healthy adults.

Measures used.

Embedding. The relative difficulty of embedded versus non-embedded clauses has been in the focus of much psycholinguistic research for decades, see for instance Lewis (1996) or Gibson (1998) on the problem of centre-embedding. Saffran, Berndt and Schwartz (1989) for instance use embedding as one indicator for structural complexity to assess the language production abilities of speakers with aphasia; Rosenberg and Abbeduto (1987) employ a similar criterion to assess production in children.

The processing of structures with embedded clauses can be argued to be more difficult on semantic and syntactic grounds: The conceptual structure of such sentences is more complex, since the embedding relation must be encoded propositionally (e. g. Bloom et al., 1980; Schleppegrell, 1992; Yuasa & Sadock, 2002). On the syntactic level, formal relationships between matrix and embedded clause need to be represented, where different types of embedding (e. g. complement, adjunct, or relative clauses) might incur different processing difficulty, depending on the kind of relation that has to be processed (cf. Speer & Clifton, 1998; Schütze & Gibson, 1999; but also see Przepiórkowski, 1999), and depending on the depth of embedding (Delage & Tuller, 2010). For these reasons, we searched for tokens of embedded structure in our data.

Word order/canonicity. Other markers that have been argued to be indicative of structural complexity are the word order differences that distinguish for instance SVO and OVS sentences in German or other V2 languages, active and passive sentences, or subject relative clauses from object relative clauses. According to for instance Fanselow et al. (1999), what is particularly costly in 'non-canonical' sentences is the dislocation of arguments and the concomitant difficulty in assigning the right theta-role to the right argument. Gorrell (2000) presents evidence for a subject-before-object preference in German, which is violated in sentences with a topicalised object or in passives (but see Drai & Grodzinsky, 2006 for discussion of the status of passive sentences in German). In addition, information structure and discourse context have been shown to play a role in how arguments are linearised and how much difficulty a particular linearisation poses during processing (Weskott, 2002; Späth, 2003; Mak, Vonk & Schriefers, 2008). Nevertheless, without context the canonical word order can be considered 'easier' to process than a non-canonical word order. Based on this premise we used the amount of instances of non-canonicity as one of our dependent measures.

Other measures. In addition to the structural indicators we counted the occurrence of grammatical errors and calculated measures for the Lombard reflex as well as for speech rate and fluency, in order to replicate effects from earlier studies and establish whether the background noise we used would have an observable effect on these aspects of language production.

Spontaneous production. For the current study, we elicited semi-spontaneous speech samples from speakers under different conditions. Using spontaneous speech or conversation recordings has a couple of advantages over more strictly controlled methods for testing language production (cf. Eisenbeiss, 2010). First and foremost the ecological validity of studies using naturalistic data is typically higher, because of the reduced artificiality of the task, compared to other elicitation methods. In addition, typically no (or much fewer) preconceived theoretical considerations can influence the design of stimuli. Of course, for the very same reasons, the conclusions that can be drawn from naturalistic or structured sampling are inherently limited, due to the lack of systematic experimental control. Samples may vary greatly between subjects in size and content, and hence the natural contexts they might provide for particular phenomena can vary too (Eisenbeiss, 2010).

Despite these issues, naturalistic speech samples provide a good starting point for explorative research and hypothesis generation (Brown & Hanlon, 2004). For the current study we compare speakers' performance under different noise conditions within subject, and we do not perform comparisons between groups that systematically differ in properties inherent to the subjects. Therefore we should be able to interpret possible effects of noise causally, despite the limitations of the data collection method described before.

Expectations. Based on the few earlier studies about sentence production in noise, we expect changes in voice intensity and/or quality (Lombard effect), speech and error rate, as well as a potential reduction in the amount of complex or difficult sentence structure. Because of its exploratory nature, the current study is intended as one of the first steps to investigate the influence of a detrimental, noisy communication setting on syntax in language production. The research presented here might serve as a baseline for future studies testing other populations.

METHOD

Participants

We tested 12 subjects, 7 female, between 20 and 30 years of age (mean = 23.4; SD = 3.15). Subjects were recruited from among students of the University of Oldenburg and the Oldenburg Technical College, and they were paid 10 Euro per hour for participation. All participants reported to have no known history of language or speech disorders, or hearing impairments.

In order to estimate the role of differences between participants in terms of their cognitive capacities, we tested each subject with a *reading span* task (Daneman & Carpenter, 1980). This kind of test has been argued to be sensitive to individual differences in verbal working memory, especially with respect to tasks which put both storage as well as processing demands on language users (Just & Carpenter, 1992, but see Waters & Caplan, 1996). We tested subjects with German sentences modelled after the material used by Daneman and Carpenter (1980), and presented sentences in random order, as suggested by Friedman and Miyake (2005). A weighted score was obtained for each participant, taking into account the amount of correctly remembered sentence-final words per trial.

Material

Three kinds of stimulus material were used to elicit spoken language from participants, two sets of interview questions, two short, one-page picture stories, and a 24-page long picture story. The interview questions revolved around the two topics holidays/festive days, and travel (see appendix). In addition, we selected two one-page picture stories from the *Father and Son* set of stories by German cartoonist e. o. plauen (Ohser, 1982). Both stories are made up of six black and white drawings without text, which depict individual scenes that connect in a self-contained plot. Finally, the long picture story we used was one of the text-free *Frog Stories* by Mercer Mayer ("Frog, where are you?", Mayer, 1969), which we slightly shortened to fit 24 (2 × 12) single pages.

As distractor noise we used a speech-free sound signal that had been designed to model acoustic characteristics of six concurrent speakers (*ICRA 7*; Dreschler, Verschuure, Ludvigsen & Westermann, 2001). The signal possesses a frequency spectrum similar to that of speech. In addition, the intensity of the signal fluctuates (pseudo-)randomly over time, in order to imitate the 'babble' impression experienced in situations where many speakers talk at the same time. The noise signal is part of a standardised set of signals for audiological testing procedures, commissioned by the *International Collegium of Rehabilitative Audiology*.

Apparatus

Subjects were recorded in a sound-attenuated booth at the Speech and Music Lab of Oldenburg university. Recordings were made directly to hard disk, using an AKG C-1000S microphone and an Echo Audio GINA 3G low latency sound adapter. The software used for recordings was PRAAT (Boersma & Weenink, 2010). Sound files containing the noise signal were played back through two Genelec 8020A loudspeakers, facing the subject at approximately 110 cm distance. In order to perform intensity measurements on the sound recordings, we calibrated the hard- and software setup using a Brüel & Kjær Investigator 2260 sound pressure level (SPL) meter. Root mean square (RMS) sound pressure level of the noise alone was 65 dB SPL at microphone position. Before every recording, the setup was checked for correct position of microphone and loudspeakers.

Procedure

Subjects were tested individually. After briefing, a reading span score was determined for each subject. The recording session for the elicitation consisted of three parts, (i) a semi-standardised interview, (ii) two short picture story descriptions, and (iii) a long picture story description task. Each part was again divided into two halves, one half of the task carried out in silence, the other half in noise. The order of the two noise conditions was counterbalanced across subjects and tasks so that each half of each tasks was carried out under noise or in silence equally often. An entire study session lasted about 45 to 60 minutes per subject, with an average net recording time of 15 minutes per subject (ranging between 10 and 32 minutes).

The interview was carried out by the experimenter, who sat opposite of the subject at a table inside the sound-attenuated booth while asking the interview questions. For the second task, each of the two short picture stories was presented on a single sheet of paper, and subjects were allowed to look at the entire story before starting to speak. The instruction given to the subjects asked them to narrate the story, rather than only describing the pictures' content. The final task of the participants was to recount the wordless picture story: "Frog, where are you?" (Mayer, 1969). In order to be able to record under the two noise conditions, the story was divided into two halves of 12 pages each. Each picture was presented on an individual page, and subjects turned the pages themselves. In this case subjects were not allowed to look ahead before turning the page. As with the short picture stories, the instruction was to narrate the events, rather than describing the picture contents.

Scoring and analysis

Sound recordings were transcribed by a trained transcriber, using the CHAT transcription standard (MacWhinney, 2000), and were verified by the first author. Utterance boundaries were determined according to criteria given in the CHAT manual, based on prosodic, syntactic and semantic features, with priority of the former two types of features over the latter.

Since the recordings for each task differed in length between subjects, a sample was taken from each transcript that was approximately 300 (interviews), 150 (short picture story), or 400 (long picture story) words long for each half of the three tasks. Samples were drawn with the *kwal* command of the CLAN tools (MacWhinney, 2000) from the middle of each transcript and were extended to start and finish at utterance boundaries. For the acoustic analyses, the audio recordings were manually trimmed at the sample boundaries determined for the transcripts. For the measurements of intensity and speech rate we used the samples from our recordings, for which some additional signal processing and cleansing was necessary.

Speech rate. In order to obtain a speech rate measure for each sample file, we manually edited the audio files to remove experimenter speech from the interview recordings. We counted the amount of words spoken by the subjects in a sample using CLAN, and calculated speech rate in *words per minute* (WPM) based on a measurement of the 'cleaned' sample audio file duration. For the speech rate measure, we kept all pauses made by the subject, since they obviously contribute to the speech rate construct.

Intensity. In order to test for an increase in vocal intensity, indicative of a Lombard effect, we calculated the root mean square intensity of speech in both noise and silence. Since the calculation of a mean intensity value might be skewed by potential differences in the amount and length of pauses the subjects made in the two different conditions, we decided to remove all speech-free portions longer than 0.7 sec (in addition to experimenter speech). The search for pauses was done with the help of a script for the PRAAT software by Lennes (2006). The results of the automatic search were manually verified, and the marked pauses were removed from the audio files. The RMS sound pressure in Pascal (Pa) was obtained for each processed sample audio file using PRAAT.

For the recordings we had opted for a free-field study setting, with noise exposure through loudspeakers. We considered this presentation method to be more natural than requiring subjects to wear headphones. For this reason, the sound recordings of the subjects' speech also contain a noise signal portion. In order to estimate the 'pure' speech SPL in noise L_s, we measured the RMS sound pressure p_n of a calibration recording that only contained noise and subtracted this value from the RMS sound pressure p_{sn} we had measured for each recording of speech and noise combined. Accordingly, for recordings of speech in silence we subtracted an RMS sound pressure p_n that was obtained from a silence recording with the same setup:

$$L_s = 10 \cdot \log\left(\frac{p_{sn}^2 - p_n^2}{p_{ref}^2}\right) \mathrm{dB}$$

p_s, p_{sn} and p_{ref} were measured in Pascal (Pa), p_{ref} is a reference value for calculating the sound pressure *level* of a sound event in decibels (dB), and is conventionally assumed to be $2 \cdot 10^{-5}$ Pa for sound travelling air.

Complexity measures. Further analyses involving structure counts were carried out manually by the first author, using the transcripts. A number of different indicators for structural complexity were obtained from the transcribed samples. Tokens of complex structure types according to the criteria given earlier, counted as indicators for complexity in production of subjects.

The value for *mean clauses per utterance* (MCU) takes into account the total number of main clauses and all embedded clauses in an utterance (e. g. Kemper, Kynette, Rash, O'Brien & Sprott, 1989; Nippold, Hesketh, Duthie & Mansfield, 2005), while a separate count of *embedded structures* contains the amount of dependent clauses only. In addition, we separately looked at three categories of structures from the embedded clauses: *relative clauses, complement clauses,* and *adverbial clauses* (cf. Hamann et al., 2007; Delage & Tuller, 2010). The effects of noise on the canonicity of sentence structures in the produced language was assessed by counting two structures we believed to model cases of *non-canonical* structure: passive sentences and sentences with a topicalised (fronted) object. Finally, for the number of *ungrammatical structures* all clauses which contained grammatical errors (e. g. lexical errors, morphological errors, missing elements, or interrupted sentences) were counted.

Design and statistical analysis. The presentation of noise was counterbalanced across the different tasks and task halves, and was a within-subject factor. Results were analysed using mixed effects models (Baayen, Davidson & Bates, 2008; Baayen, 2008; Jaeger, 2008). To analyse the different complexity measures, we calculated proportions (e. g. of embedded clauses relative to the total number of clauses), and computed linear mixed effects models with noise condition as predictor, using the *lmer* package for the R statistical software. Random effects for subjects and task (interview, short picture story, long picture story) were estimated by adding random intercepts to the models. Intensity and speech rate measures were also evaluated with linear mixed effects models. P-values for the individual model parameters were calculated with the *pvals.fnc()* function (Baayen et al., 2008).

In addition to estimating the effect of noise on subjects speaking, we entered their respective score on the *reading span task* into the model, in order to check for correlations between each outcome variables and our measure for individual working memory capacity.[3] As an additional step to analyse potential correlations with reading span, we performed a model comparison procedure as suggested by Baayen et al. (2008), in order to check whether the reading span measure would improve the fit of the respective model for each outcome variable.

RESULTS

Figures 1 through 3 show the results of our different measures in silence and noise respectively. The noise seems to have an effect on the vocal intensity with which our subjects spoke, and the amount of ungrammatical structures was significantly higher in noise than in silence. While there appear to be differences between the silence and the noise condition for other measures like MCU or the amount of non-canonical structures, these results were not statistically significant.

No correlation of subjects' reading span score with any of the different outcome variables reached significance. Table 1 provides an overview of the parameter estimates and p-values for *reading span* score as a coefficient in the different models. As our model comparisons for each outcome variable showed, adding the reading span score to the different models did not improve model fit in any case. Therefore, we removed this parameter from our model specifications. The parameter estimates for the effect of noise on the different outcome variables given in the following results are based on the simplified models.

Table 1 about here

Panel A of figure 1 shows the average *speech rate* of subjects in noise and in silence. We do not find a statistically significant difference in speech rate between silence and noise (N = 72, log-likelihood = -28.47; Coef. = -0.002, SE = 0.07, p = .980). Panel B of figure 1 shows the result of the *intensity measurement* in silence and under noise. The effect of noise reaches significance: in noise, spoken language is about 8 dB louder than in silence (N = 72, log-likelihood = -148.3; Coef. = 7.044, SE = 0.35, p <.001).

Figure 1 about here

The average number of *clauses per utterance* (MCU) for the two noise conditions is given in panel C of figure 1. The MCU appears to be slightly higher in noise than in silence, but this difference does not reach significance (N = 72, log-likelihood = -2.728; Coef. = 0.098, SE = 0.05, p = .071). Panel A of figure 2 shows the proportion of *ungrammatical clauses* out of all clauses produced by a subject for a given task, broken down by silence and noise. Subjects seem to produce slightly more ungrammatical structures under noise than in silence, which is indicated by a significant coefficient for the factor noise (N = 72, log-likelihood = 153.4; Coef. = 0.014, SE = 0.01, t = 2.48, p <.05). The grammaticality errors we observed manifest themselves in different ways. We categorised the different errors *post hoc* in order to check for noticeable patterns (see table 2 for an overview).

Figure 2 about here

Table 2 about here

The proportion of *non-canonical structures* produced in silence and in noise is presented in panel C of figure 2. We do not find an indication that noise has an effect on the amount of non-canonical structures (N = 72, log-likelihood = 86.35; Coef. = -0.018, SE = 0.02, p = .275). Panel B of figure 2 shows the proportion of *embedded clauses* produced in silence and under noise. Our statistical model (N = 72, log-likelihood = 55.36) does not give us reason to assume an effect of noise on the amount of embeddings subjects produced (Coef. = 0.036, SE = 0.02, p = .126). For more fine-grained analyses, we broke down the total amount of embedded structures, to take a separate look at different kinds of clauses: relative clauses, complement clauses, and adverbial clauses (see figure 3). However, the statistical analyses did not yield significant differences between the silence and the noise condition for either of the three categories (all p >.1).

Figure 3 about here

DISCUSSION

Reading span

The results of our study yield no indication for a correlation between reading span and any of the complexity or fluency measures. This is in contrast to earlier studies, for authors such as Miyake, Carpenter and Just (1994) or Kemper and Sumner (2001) reported such correlations and have argued for reading span to be indicative of working memory capacity, which in turn they regard as an important factor for creating and maintaining syntactic dependencies within sentences, one of the hallmarks of complex sentences. However, the validity of the reading span task as a measure for syntactic processing has been criticised by other authors, for instance Waters and Caplan (1996) who claim the reading span task is too different from processing for sentence comprehension (also see Friedman & Miyake, 2004). The lack of a significant correlation between reading span scores and our linguistic measurements might support the latter point of view, although it should be borne in mind that our observations were made on language production rather than comprehension. Given the divergent theoretical claims and empirical findings on this issue, we have to refrain from further speculation about why we did not find a correlation.

Lombard effect

That the noise condition we created has an effect on speakers is evidenced by the increase in speech intensity under noise. These results indicate that the current study successfully replicated the *Lombard reflex* or *effect*, in line with findings by for instance Lu and Cooke (2008) and a number of earlier studies on the effect. Lu and Cooke (2008) assume an automatic regulation mechanism underlying the changes in speakers voice under noisy conditions. In order for a Lombard reflex to occur, a regulatory or control loop needs access to the auditory input coming in through the perceptual system. The characteristics of this input then need to trigger changes in the articulatory system. This mechanism should in principle be compatible with control mechanisms postulated for speech monitoring, or might even form part of the monitor.

In anticipation of concerns about the study setting, we have to issue a word of caution about our measurement method: since we estimated the voice intensity of speakers from a sound recording which contained both the speaker's voice and the distractor noise, the data might be subject to a some measurement error. But given the strong experimental control of hard- and software settings for the recordings, we are confident that the observed increase reflects an actual effect of noise. Our trust in the data is supported by the fact that the magnitude of the effect we observe (approximately 8 dB) is similar to an earlier observation reported by Garnier, Bailly, Dohen, Welby and Lœvenbruck (2006), who found an increase of 8.6 dB for utterances produced in (fluctuating) 'cocktail party' noise.

Speech rate

Different from what we would have expected based on earlier studies, we did not see an effect of noise on the rate at which subjects articulate. In the light of results by for instance Postma and Noordanus (1996), or our own results from a different set of experiments (Hanke et al., in press; Hanke, Hamann & Ruigendijk, 2012), we currently have no explanation for our findings. It is possible that the method we used for measuring speech rate comes with too broad a margin of error for obtaining a significant effect, either because of the necessary sample 'cleaning' steps or because a measurement in words per minute is too coarse in scale.

Structural measures

Our different measures of structural complexity do not indicate a systematic decrease of complexity under noise. This is in contrast to earlier findings by Kemper et al. (2003), and for now we can only speculate about the reasons for the difference in outcome.

An obvious reason might be differences in terms of the experimental setting, for instance the type of noise that was used, and the intensity of presentation: The speech and noise distractor sounds used by Kemper and colleagues were presented at an intensity of 40-60 dB. Despite the lack of further details on their experimental setting, this should amount to a lower noise signal intensity than the intensity level used in our study.

One other potential factor might be that the distractor noise used by Kemper et al. (2003) contained speech or speech fragments, very likely even in the "cafeteria noise" recording. In this case, the distractor stimulus would contain linguistic material that might be more difficult to ignore than speech-free noise. Following the logic of dual-task experiments, if linguistic content is present, a competition between the processing of language for the (primary) speaking task and unconscious processing of speech in the distractor might take place, resulting in capacity limitations on shared processes. Hence, we might speculate that the amount of 'overlap' between processing resources necessary to automatically process the noise signal and linguistic processing-for-production might not be high enough to result in an observable performance decrease, when speech-free noise is used.

Additionally, the design of our study might not have resulted in a conversational setting formal enough to warrant a speaking style that is characterised by a large amount of complex structure to begin with (cf. for instance Nippold et al., 2005). It is difficult to assess, however, in how far differences between studies in terms of lab setting or the wording of instructions might be able to explain our results.

In sum, we cannot find evidence for a 'processing capacity' effect of noise on speaking with respect to grammatical complexity, contrary to what was observed by Kemper et al. (2003). We might only speculate that the amount of distraction we offered to our participants was not strong enough, compared to the distractor stimuli used by Kemper and colleagues, even given the higher intensity level at which our noise signal was presented.

Error count

The lack of conclusive results about sentence structure complexity non-withstanding, noise did affect the speech production of subjects in our study: The amount of grammaticality errors we observed was increased ever so slightly, but nonetheless significantly under noise. This is in line with earlier findings by Postma and Noordanus (1996), and complements their observations about *phonological* errors with errors on the morphological, lexical and syntactic level. The authors report a decrease in self-reported phonological error rates while producing tongue-twisters when overt, auditory feedback was suppressed for instance by (white, i. e. temporally unmodulated) noise. The noise signal was presented via headphones at 100 dB SPL, an intensity level which was certain to almost completely block external feedback.

Postma and Noordanus (1996) attribute their finding of an increased error rate to the lack of an additional monitoring channel through the (external) auditory feedback loop.

Based on Levelt's perceptual loop theory, two mechanisms seem plausible to explain our error effects, both not mutually exclusive: First, some amount of acoustic masking of the signal that is fed back through the external loop will take place, and the monitoring of one's own overt speech is less effective. Second, we might assume an effect of noise on the internal monitoring loop: While the language perception system is partially occupied with (automatic) processing of the distractor noise, the concurrent processing of internal loop information becomes impaired to the effect that erroneously specified parts of speech will be less likely to be intercepted before articulation. This kind of competition is very much akin to what is assumed to happen in studies on the *irrelevant sound effect*.[4]

Thus, fluctuating noise could impair error monitoring on two levels: acoustically, by masking/degrading the 'external' memory representation of what had been said, and cognitively, by generating competition for processing time/space on resources shared between the automatic parsing of the distractor noise and monitoring. If the self-perception through different monitoring loops is impaired, the fact that we observe an increase in errors like wrong agreement marking or interleaved sentences (cf. table 2) could actually be seen as the effect of a less efficient access to parts of the conceptual, grammatical or phonological structure of what one was about to say.[5] More generally, the 'self-perception' during speaking through the external and internal loop might actually even serve a function beyond monitoring, to reinforce the memory representation of parts of speech that have been produced earlier. Based on the sentence parsing architecture proposed by Lewis and Vasishth (2005), Badecker and Kuminiak (2007) have suggested that the on-line generation of structural dependencies like number agreement crucially relies on reactivation of parts of the earlier utterance from immediate memory.

The Lombard reflex might be different in terms of the regulatory mechanism involved, which influences muscular control necessary for phonation and articulation, but in effect the reflex might even be in a speaker's own interest: to reinforce the self-perception during speaking, in order to reduce the likelihood of errors or to even serve memory retrieval processes necessary for the construction of sentence structure on-the-fly.

CONCLUSION

While the literature yields some evidence for performance decreases in acoustically difficult situations, healthy adult speakers appear to be sufficiently fine-tuned to perform the task of speaking even in a difficult, noisy environment. Our results indicate that there is little overlap between 'low-level' acoustic processing of *speech-free* noise and formulation processes, with the exception of monitoring for grammaticality errors. The monitor might be impaired by competition for processing resources needed to access earlier parts of the spoken utterance when monitoring through the internal loop, and in addition to that detriments might arise by energetic masking of the self-perception through the external loop. Self-perception could serve to provide or strengthen search cues to re-access earlier parts of the utterance from immediate memory, at positions when this is necessary to process the current part of the utterance. In order to further investigate the effects of noise on language production, more strictly controlled experimental tasks should follow, which increase the burden on processing capacity. In addition, such experiments would allow for a closer examination of the effects of different kinds of noise, especially more speech-like ('babble') noise types.

NOTES

1. A couple of other loops have been proposed since by other authors. Postma (2000) gives an overview and cites evidence for as many as 11 monitoring pathways within the language production system.
2. This is one of the reasons why MLU is considered to work as an estimate of grammatical abilities only up to a certain age (Brown, 1973; Dethorne, Johnson & Loeb, 2005); however see Szmrecsányi (2004).

3. While both parameters are estimated in one model, our interpretation of the two predictors used in the model differs in that we attempt causal inference from the noise factor, which was controlled and within-subject, whereas for the correlation between results on the reading span task and a particular result on our outcome variable we will not attempt a causal explanation in this study (Shadish, Cook & Campbell, 2002).
4. It is possible, of course, that a competition for processing resources instigated by irrelevant sound also affects the effectiveness of the external monitoring loop, in addition to the degradation of the perceived signal. A seriation mechanism as proposed by (Jones, 1993) that keeps track of the order of information objects stored in memory could be a sub-component of the monitoring system.
5. The question remains, however, how impaired monitoring through the *external* loop, which always happens after articulation has already taken place, can lead to a higher number of grammaticality errors. Since our detailed error analysis was carried out *post hoc* and because the types of errors found in our study do not preclude an explanation based on external monitoring, we have to leave this question open for now.
6. We featured the three examples as 'major' holidays since they involve official bank holidays all over Germany, and are usually accompanied by at least one week of school vacation in most parts of the country.

REFERENCES

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Badecker, W. & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language, 56,* 65–85.

Banbury, S. & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology, 89,* 499–517.

Banbury, S. P., Macken, W. J., Tremblay, S. & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors, 43*(1), 12–29.

Bloom, L., Lahey, M., Hood, L., Lifter, K. & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language, 7,* 235–261.

Boersma, P. & Weenink, D. (2010). *Praat: Doing phonetics by computer* (Version 5.2.06) [Computer software]. Retrieved 20 December 2010 from http://www.praat.org

Brannon, J. B. & Murry, T. (1966). Spoken syntax of normal hard-of-hearing and deaf children. *Journal of Speech and Hearing Research, 9*(4), 604–610.

Brown, R. W. (1973). *A first language: The early stages.* London: Allen & Unwin.

Brown, R. W. & Hanlon, C. (2004). Derivational complexity and order of acquisition in child speech. In B. Lust & C. Foley (Eds.), *First language acquisition: The essential readings* (pp. 155–175). Malden, MA: Blackwell. (Reprinted from *Cognition and the development of language*, by J. R. Hayes, Ed., 1970, New York: Wiley)

Carroll, R. & Ruigendijk, E. (in press). The effects of syntactic complexity on processing sentences in noise. *Journal of Psycholinguistic Research.* doi: 10.1007/s10936-012-9213-7

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466.

Delage, H., Monjauze, C., Hamann, C. & Tuller, L. (2008). Relative clauses in atypical acquisition of French. In A. Gavarró & M. J. Freitas (Eds.), *Language acquisition and development: Proceedings of GALA 2007.* Cambridge: Cambridge Scholars Publishing.

Delage, H. & Tuller, L. (2010). Evolution of syntactic complexity and avoidance strategies in children and adolescents with mild-to-moderate hearing loss. In J. Costa, A. Castro, M. Lobo & F. Pratas (Eds.), *Language acquisition and development: Proceedings of GALA 2009.* Cambridge: Cambridge Scholars Publishing.

Dethorne, L. S., Johnson, B. W. & Loeb, J. W. (2005). A closer look at MLU: What does it really measure? *Clinical Linguistics & Phonetics, 19*(8), 635–648.

Drai, D. & Grodzinsky, Y. (2006). The variability debate: More statistics, more linguistics. *Brain and Language, 96,* 157–170.

Dreschler, W. A., Verschuure, H., Ludvigsen, C. & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *Audiology, 40*(3), 148–157.

Eisenbeiss, S. (2010). Production methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 11–34). Amsterdam: John Benjamins.

Ellermeier, W. & Hellbrück, J. (1998). Is level irrelevant in “irrelevant speech”? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1406–1414.

Fanselow, G., Kliegl, R. & Schlesewsky, M. (1999). Processing difficulty and principles of grammar. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar, and memory* (pp. 171–201). Dordrecht: Kluwer.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *The psychology of reading* (pp. 559–586). Hove: Erlbaum.

Friedman, N. P. & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*(1), 136—151.

Friedman, N. P. & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, 37*(4), 581–590.

Friedmann, N., Belletti, A. & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua, 119*(1), 67–88.

Friedmann, N. & Szterman, R. (2006). Syntactic movement in orally trained children with hearing impairment. *Journal of Deaf Studies and Deaf Education, 11*(1), 56–75.

Garnier, M., Bailly, L., Dohen, M., Welby, P. & Lœvenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. In *Proceedings of the 9th International Conference on Spoken Language Processing INTERSPEECH 2006 – ICSLP*. Pittsburgh, PA: International Speech Communication Association.

Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Speech and talk* (Vol. 1, pp. 177–220). London: Academic Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68,* 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gorrell, P. (2000). The subject-before-object preference in German clauses. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 25–63). Dordrecht: Kluwer.

Hamann, C., Tuller, L., Monjauze, C., Delage, H. & Henry, C. (2007). (Un)successful subordination in French-speaking children and adolescents with SLI. In *Proceedings of the 31st Boston University Conference on Language Development*. Boston, MA.

Hanke, M., Hamann, C. & Ruigendijk, E. (in press). On the laws of attraction at cocktail parties: Babble noise influences the production of number agreement. *Language and Cognitive Processes.*

Hanke, M., Hamann, C. & Ruigendijk, E. (2012). *Effects of noise and hearing impairment on the production of complex sentence structure.* Manuscript in preparation.

Hanley, T. D. & Steer, M. D. (1949). Effect of level of distracting noise upon speaking rate, duration and intensity. *Journal of Speech and Hearing Disorders, 14,* 363–368.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.

Jakubowicz, C. (2010). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua, 121,* 339–351.

Jones, D. M. (1993). Objects, streams, and threads of auditory attention. In A. D. Baddeley, D. E. Broadbent & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control* (pp. 87–104). Oxford: Oxford University Press.

Jones, D. M. & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 369–381.

Junqua, J.-C., Fincke, S. & Field, K. (1999). The Lombard effect: A reflex to better communicate with others in noise. In *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 4, pp. 2083–2086). Phoenix, AZ: IEEE.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.

Kemper, S., Herman, R. E. & Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging, 18*(2), 181–192.

Kemper, S. & Kemtes, K. A. (1999). Limitations on syntactic processing. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar, and memory* (pp. 79–105). Dordrecht: Kluwer.

Kemper, S., Kynette, D., Rash, S., O'Brien, K. & Sprott, R. (1989). Life-span changes to adult's language: Effects of memory and genre. *Applied Psycholinguistics, 10*(1), 49–66.

Kemper, S. & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging, 16*(2), 312–322.

Klatte, M. & Hellbrück, J. (1993). Der "Irrelevant Speech Effect": Wirkungen von Hintergrundschall auf das Arbeitsgedächtnis [The "irrelevant speech effect": Effects of background noise on working memory]. *Zeitschrift für Lärmbekämpfung, 40,* 91–98.

Klatte, M., Kilcher, H. & Hellbrück, J. (1995). Wirkungen der zeitlichen Struktur von Hintergrundschall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen [Effects of temporal structure of background noise on working memory and their theoretical and practical implications]. *Zeitschrift für Experimentelle Psychologie, 42*(4), 517–544.

Kolk, H. (1995). A time-based approach to agrammatic production. *Brain and Language, 50*(3), 282–303.

Lennes, M. (2006). *Pause detection script (Version 23.1.2006)* [Computer software]. Helsinki: Author. Retrieved 1 August 2011 from http://www.helsinki.fi/~lennes/praat-scripts/public/mark_pauses.praat

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research, 25*(1), 93–115.

Lewis, R. L. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29,* 375–419.

Lombard, É. (1911). Le signe de l'elevation de la voix [The sign of the elevation of the voice]. *Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx, 34,* 101–119.

Lu, Y. Y. & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *Journal of the Acoustical Society of America, 124*(5), 3261–3275. doi: 10.1121/1.2990705

Macken, W., Tremblay, S., Alford, D. & Jones, D. (1999). Attentional selectivity in short-term memory: Similarity of process, not similarity of content, determines disruption. *International Journal of Psychology, 34*(5-6), 322–327.

MacWhinney, B. (2000). *CHILDES: Tools for analyzing talk.* Mahwah, NJ: Lawrence Erlbaum.

Mak, W. M., Vonk, W. & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory & Cognition, 36*(1), 170–81.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Press.

Miyake, A., Carpenter, P. A. & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology, 11*(6), 671–717.

Nippold, M. A., Hesketh, L. J., Duthie, J. K. & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research, 48,* 1048–1064. doi: 10.1044/1092-4388(2005/073)

Ohser, E. (1982). *Vater und Sohn [Father and son]* (D. Laubach, Ed.). Konstanz: Südverlag.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220–244.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*(2), 97–131.

Postma, A. & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39,* 375–392.

Przepiórkowski, A. (1999). *Case assignment and the complement/adjunct dichotomy: A non-configurational constraint-based Approach.* Unpublished doctoral dissertation, Universität Tübingen, Tübingen, Germany. Retrieved 14 April 2012 from http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-722

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology, 20*(3), 241–248.

Rabbitt, P. M. A. (1991). Mild hearing-loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica,* Suppl. 476, 167–176.

Roark, B., Mitchell, M. & Hollingshead, K. (2007). Syntactic complexity measures for detecting mild cognitive impairment. In *Proceedings of the 2007 BioNLP workshop: Biological, translational, and clinical language processing,* Prague, Czech Republic.

Rosenberg, S. & Abbeduto, L. (1987). Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics, 8*(1), 19–32.

Saffran, E. M., Berndt, R. S. & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37,* 440–479.

Salamé, P. & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 150–164.

Schleppegrell, M. J. (1992). Subordination and linguistic complexity. *Discourse Processes, 15*(1), 117–131.

Schütze, C. T. & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language, 40*(3), 409–431. doi: 10.1006/jmla.1998.2619

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston, MA: Wadsworth Publishing.

Speer, S. R. & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition, 26*(5), 965–978. doi: 10.3758/BF03201177

Späth, A. (2003). The linearization of argument DPs and its semantic reflection. In H. Härtl & H. Tappe (Eds.), *Mediating between concepts and grammar* (pp. 165–195). Berlin: Mouton de Gruyter.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B. & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science, 11*(2), 153–158.

Szmrecsányi, B. (2004). On operationalizing syntactic complexity. In G. Purnelle, C. Fairon & A. Dister (Eds.), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis* (Vol. 2, p. 1032-1039). Louvain-la-Neuve: Presses universitaires de Louvain.

Uslar, V., Ruigendijk, E., Hamann, C., Brand, T. & Kollmeier, B. (2011). How does linguistic complexity influence intelligibility in a German audiometric sentence intelligibility test? *International Journal of Audiology, 50,* 621–631.

Waters, G. S. & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*(4), 761–772.

Weskott, T. (2002). *Information structure as a processing guide: The left periphery of German verb-second sentences and its interpretation in context.* Unpublished doctoral dissertation, Universität Leipzig, Leipzig, Germany. Retrieved 14 April 2012 from http://www.sfb632.uni-potsdam.de/~weskott/diss.pdf

Wingfield, A., Peelle, J. E. & Grossman, M. (2003). Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *Aging, Neuropsychology, and Cognition, 10*(4), 310–322.

Wingfield, A. & Tun, P. A. (1999). Working memory and spoken language comprehension: The case for age stability in conceptual short-term memory. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, grammar, and memory* (pp. 29–52). Dordrecht: Kluwer.

Yuasa, E. & Sadock, J. M. (2002). Pseudo-subordination: A mismatch between syntax and semantics. *Journal of Linguistics, 38,* 87–111.

APPENDIX

Interview questions: Holidays

In this part of the interview I would like to talk with you about the 'major' holidays, like Easter, Pentecost, or Christmas.[6]

1. Can you remember the last big holiday occasion(s)?
2. Do you and your family have any particular rituals for celebrating larger holidays?
3. Are there special meals you would eat on one of those occasions?
4. Do you experience a lot of hustle and bustle over the holidays?
5. How would you cope with holiday stress?
6. What was the last present you got for a holiday occasion?
7. If tomorrow were Christmas again and you could make a wish for something, what would that be?

Interview questions: Travel

In this part of the interview, I would like to talk to you about holiday travel.

1. The last time you went on holiday, where did you travel?
2. How did you travel there?
3. Why did you travel by *[helicopter]*?
4. What is your favourite means of transport for going on holiday?
5. What do you like about travelling by *[helicopter]*?
6. How do you usually prepare for a trip?
7. If you could travel anywhere, where would you like to go?
8. Why would you like to go to *[Mars]*?

Tables

Table 1: Correlations between weighted reading span measure and different outcome variables: coefficients for reading span as parameter in linear mixed effects models.

Outcome variable	Coef. β	SE β	t	**p**
Speech rate	−0.406	0.88	−0.46	.782
Intensity (dB SPL)	4.833	3.96	1.22	.089
MCU	−0.136	0.29	−0.47	.615
Error count	−0.038	0.03	−1.39	.138
Non-canonical structures	−0.030	0.06	−0.52	.632
Embedded structures	−0.157	0.10	−1.53	.122

Table 2: Error counts in silence and noise, broken down by error type.

Error type	silence	noise
lexical errors	3	2
missing word	1	7
case errors	1	1
person, number, gender	–	4
word order	3	1
interleaved sentences	–	3
other	–	3

Figure titles

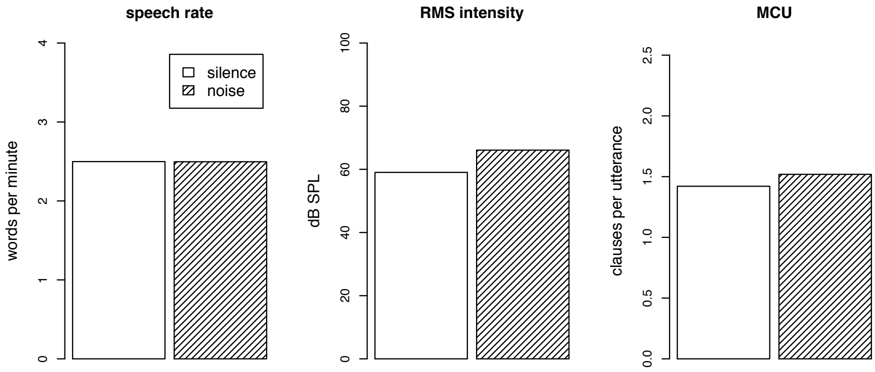
Figure 1: Speech rate (panel A), root mean square (RMS) intensity (panel B), and mean number of clauses per utterance (MCU; panel C), for silent and noise conditions.

Figure 2: Proportion of ungrammatical clauses (panel A), embedded clauses (panel B), and non-canonical structures (panel C) in silence and noise.

Figure 3: Proportion of embedded clauses by type: relative clauses (A), complement clauses (B), and adverbial clauses (C).

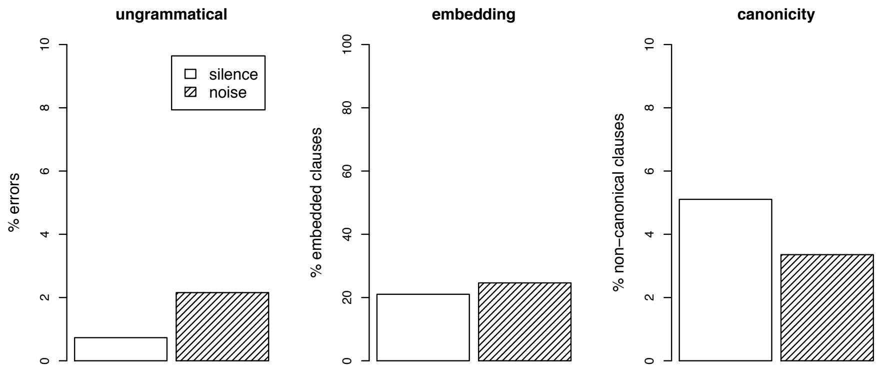
Figures

Figure 1.



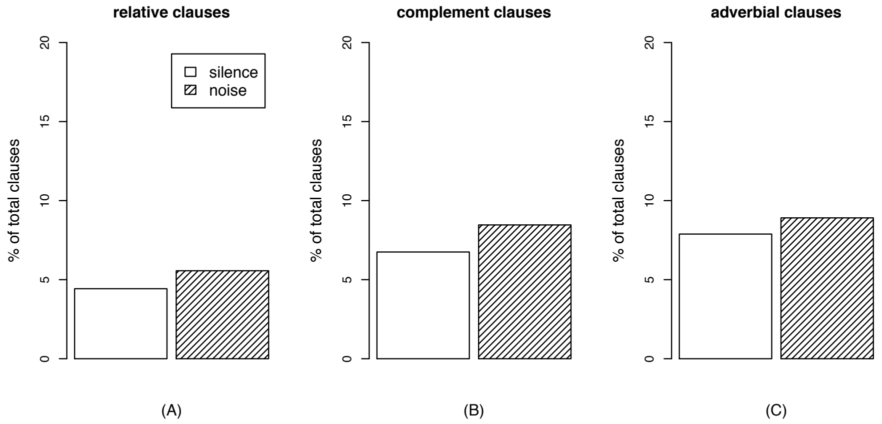
(A) (B) (C)

Figure 2.


ungrammatical
embedding
canonicity
% errors
% embedded clauses
% non−canonical clauses
0
2
4
6
8
10
0
20
40
60
80
100
0
2
4
6
8
10
☐ silence
▨ noise

(A) (B) (C)

Figure 3.


relative clauses
complement clauses
adverbial clauses
% of total clauses
0
5
10
15
20
☐ silence
▨ noise
(A)
(B)
(C)