On the laws of attraction at cocktail parties: Babble noise influences the production of number agreement

Short title: Babble noise influence on agreement

Mirko Hanke*, Cornelia Hamann, and Esther Ruigendijk
School of Language and Cultural Studies,
University of Oldenburg, 26111 Oldenburg, Germany

Please note: This is an Author's Accepted Manuscript of an article published in Language and Cognitive Processes, 28(8), 2013, copyright Taylor & Francis, available online at: http://www.tandfonline.com/10.1080/01690965.2012.696664.
(Published online in Language and Cognitive Processes iFirst*: 11 September 2012)*

**Corresponding author:*
Email mh@mirkohanke.de
~~*Phone +49 441 3090904*~~

Abstract

Theoretical accounts of the language production process have claimed that grammatical encoding steps during the formulation stage happen in a largely automatic fashion, unimpeded by other cognitive processes (cf. Meyer, Wheeldon, & Krott, 2007: x). By eliciting *agreement attraction errors*, our study tested the effect of external distractor noise on the generation of subject-verb agreement in spoken language. We modelled noisy environments with three different speech-free sounds with spectral, or both spectral and intensity modulation characteristics of speech. In silence and unmodulated noise we found evidence for a plural mismatch effect, where a plural local noun attracts agreement away from a singular marked head noun. Under modulated noise the error patterns changed, and the number of errors increased in cases where the head noun of the preamble was marked for plural. In addition, background noise led to a reduction of speech rate and a reduction of utterance latency. Our results indicate that unspecific, speech-free noise can create a secondary task load which exerts an influence on the grammatical encoding stage of sentence formulation. We suggest that additional load leads to a slowing down of processing and subsequent difficulty to access the correct number information of the head noun when trying to retrieve an inflected verb form. Subjects overcome this difficulty by resorting to default singular marking on the verb. The results of this study challenge previous claims about the informational encapsulation of the grammatical encoding stage.

Acknowledgements

We would like to thank Hendrikje Ziemann for her valuable help in running the experiments and carrying out measurements. We would also like to thank Patrick Sturt and the two anonymous reviewers at LCP for their constructive comments and suggestions which were of great help to improve the paper. The research reported in this paper was funded by the DFG German Research Foundation, grant HA 2335/2-1 awarded to the second author.

Introduction

The task of using language in *noise* has increasingly come to the attention of psycholinguists. We speak and comprehend under various different circumstances: Imagine the typical psychological or (psycho-)linguistic department and its laboratory setting: carpet on the floor to silence steps, hushed voices, and sometimes even sound-attenuated rooms in which test subjects are made to speak, or listen to language. Now imagine a large city, with cars, buses, trams, people screaming, pedestrians talking over the noise to others next to them or on the phone. Admittedly, these two examples form a rather extreme opposition, but it should be intuitively clear from these imagined settings that while our ability to speak and comprehend language has evolved to function under very different external settings, the presence of external noise imposes the feeling of greater effort in both speaking and comprehending. For comprehension in noise, an initially straightforward explanation might be that parts of the heard message are masked by noise. In contrast to that, it is far less obvious whether and how noise could influence speaking. It can be asked in particular whether the formulation stage, which has been argued to be an example of a modular, automatic process (Bock, 1982; Levelt, 1989; Ferreira, 2007), affected by external factors like noise?

Speaking in noise

Audiological and acoustic research on the so-called *cocktail party phenomenon* has emphasised the effect on comprehension of the acoustic characteristics of background noise created by multiple speakers talking (see for instance Bronkhorst, 2000). The 'buzz' or 'babble' at cocktail parties possesses particular acoustic characteristics that are different from 'static' (white or pink) noise. The temporal characteristics of such a signal can be described as showing broad-scale fluctuations in amplitude across time, creating an envelope of intensity peaks and troughs, which are partially responsible for creating the 'babble' impression.

Studies in the field of attention and memory psychology have been investigating the effects of cocktail party situations on verbally mediated cognitive tasks for quite some time. Research about the so-called *irrelevant speech or sound effect* (ISE) 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, Macken, Tremblay, & Jones (2001) for a review. Somewhat counter-intuitively, the effect is independent of intensity or content of concurrent speech (Klatte, Kilcher, & Hellbrück, 1995; Ellermeier & Hellbrück, 1998; Salamé & Baddeley, 1982). Importantly, the effects of irrelevant auditory stimuli on verbal recall are also not limited to irrelevant speech. As Jones and Macken (1993) and Klatte and Hellbrück (1993) have shown, series of intermittent tones or content-free noise lead to performance decreases in recall tasks. What is more, the negative effect on performance is substantially stronger with fluctuating, 'babble-like' noise than with constant noise (Klatte et al., 1995). What these results indicate is that simply listening to, or ignoring certain kinds of noise can create a secondary (or dual-) task load on certain cognitive tasks. But because speaking crucially involves the generation of linguistically structured material rather than recalling ordered lists of otherwise unconnected words, an obvious, yet still open question is whether noise has an impact on any of the various cognitive processes that allow us to produce language.

This issue touches directly upon the question of *cognitive modularity*: In his information processing theory of speaking, Levelt (1989) addresses the question of cognitive resources that might be involved in speaking. According to his account, an attentionally guided process like message generation (the 'conceptualisation' stage) demands resources and is relatively slow. Further steps in the process of speaking are however conceived of as largely automatic and, crucially, are assumed to have dedicated processing capacity at their disposal (Levelt, 1989: 20; also see Caplan & Waters, 1999 and Fodor, 1983), notably the formulation stage, during which lexical items are selected, grammatical relations are established and constituent structure is created (Levelt, 1989; Bock & Levelt, 1994).

However, this rather strong claim that the formulation stage operates automatically and without sharing processing resources with other cognitive functions has been called into question by other authors. For instance, a study by Ferreira and Pashler (2002) established effects of a dual-task load on some of the stages of lexical selection. In two experiments carried out by Oomen and Postma (2002), the number of errors detected in two monitoring tasks was reduced when a secondary task load was present. Fayol, Largy, and Lemaire (1994) found a higher number of agreement errors in written reproduction of memorised sentences when subjects had to perform a concurrent task, like counting or maintaining unrelated items in memory. Finally, results by Hartsuiker and Barkhuysen (2006) have shown both an effect of individual differences in working memory and an effect of concurrent memory load on the production of subject-verb agreement in Dutch. Taken together, the different studies indicate that the automaticity of processing in the formulation stage, including grammatical encoding, might be more a matter of degree, rather than an all-or-nothing property, and that formulation is not entirely 'resource-free' (cf. Hartsuiker & Barkhuysen, 2006). That is, formulation might make use of shared (verbal) working memory resources to some extent; see for instance Garrod and Pickering (2007) for a detailed discussion of this issue.

In the current study we followed this line of research and tried to establish whether external distractors, in our case different speech-free *noise* signals exert an influence on grammatical encoding. Or, in other words, we wanted to test whether the sheer presence of noise will establish a dual-task load that interferes with procedures implied in language production. We were particularly interested in the grammatical encoding stage, during which grammatical functions are assigned and constituent structure is assembled, because this processing stage is typically assumed to proceed in a largely automatic, 'modular' fashion.

So far, empirical work that has dealt with the effect of noise on speaking has been concerned with for instance acoustic or prosodic effects, in work on the so-called *Lombard reflex* or *Lombard effect*, an increase in perceived articulation effort (Lombard, 1911; see Junqua, Fincke, & Field, 1999 or Lu &

Cooke, 2008 for more recent work on this effect). Another line of research investigated the effect of noise on self-*monitoring* during speaking, by means of internal and overt feedback loops (e. g. the *perceptual loop*, using the language comprehension system; Levelt, 1983; Levelt, 1989). Postma and Noordanus (1996) tried to block the external feedback loop of speakers producing tongue-twisters by presenting them with white noise at a high intensity through headphones. The authors found a general slowing down of articulation in noise. In addition, the number of self-reported phonological speech errors was reduced. This result indicates that monitoring through the external feedback link provides an additional channel to detect errors, which can be blocked or impaired in noise. Finally, language production research targeting effects of cognitive ageing provided evidence that there might be an effect of noise on the sentence production process such that the generation of syntactic structure would be influenced. A study by Kemper, Herman, and Lian (2003) looked for detrimental effects of secondary tasks, such as listening to and ignoring cafeteria noise, on the linguistic complexity of utterances. The authors found effects of noise on spoken language performance, which were different for the two age groups compared in the study. While older speakers showed a reduction of speech rate under difficult conditions, younger adults reduced sentence length and, crucially, grammatical complexity. These results not only again indicate that ignoring irrelevant noise can generate a secondary task load, but also suggest that this load can impact the generation of linguistic structure during the grammatical encoding stage.

However, as Hartsuiker and Barkhuysen (2006) note, the measures employed by Kemper et al. (2003) might be too coarse to pinpoint a potential location of (or mechanism for) interference. Therefore, we opted to take a closer look at the production of *agreement* between subject and verb as a testing case, and we examined the influence of external noise on the number of agreement errors.

Agreement and agreement errors

Agreement in natural languages marks grammatical relations between elements of a sentence. In the case of agreement between verb and subject, the signalled relation is subjecthood, and since typically the subject bears the external argument role of a verb, the grammatical relation stands for a particular thematic or propositional configuration. This is realised as follows: an agreement *target* (e. g. the verb) bears identical values for one or more grammatical features as the so-called agreement *controller* (e. g. the subject of a sentence; Moravcsik, 2006), and this featural identity can then be signalled on a morphological level, for instance by the verb being realised in a particular form that represents a particular setting of feature values.

In German, lexical and auxiliary verbs agree with the subject noun phrase in person and number, to the effect that the verb bears an inflection ending which signals 1st, 2nd or 3rd person, and singular or plural number respectively. Nouns typically show inflection for number, with differences in terms of the morphological realisation between several noun classes. Furthermore, there is regular case, gender and number agreement between a noun and its determiner, and between adjectives and nouns. Depending on the combination of grammatical features, the formal realisation of case and number on the article and/or adjective can provide an additional cue along with the noun inflection (cf. also Hartsuiker, Schriefers, Bock, & Kikstra, 2003).

Experimental data about *agreement attraction errors* has been playing a major role in the development of psycholinguistic models for the computation of agreement for years. Different factors and conditions have been identified that proliferate the occurrence of number agreement errors as in example 1, where the number marking on the verb seems to be 'attracted' by an element other than the actual controller of verb agreement:

(1) "However, it is only the meaning of the words which **have* changed, not the grammatical structure of the language."

J. F. Schereschewsky, Hartford Courant, 5 November 1964; quoted after Francis (1986).

Quite a number of studies have exploited the attraction effect in order to investigate the processing of syntactic information in both spoken (Bock & Miller, 1991; Schriefers & van Kampen, 1993; Vigliocco & Nicol, 1998) and written language production (Fayol et al., 1994; Hölscher & Hemforth, 2000;

Hemforth & Konieczny, 2003), as well as in written language comprehension (Pearlmutter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009). Previous studies have tried to assess the role of different, mostly linguistic factors on the generation of agreement errors. Already some of the earliest descriptions of attraction errors in the linguistic literature noted an effect of *linear distance* between agreement controller and an interpolated noun on the likelihood of attraction errors, and consequently dubbed the effect 'proximity concord' (Hale & Buck, 1903/1966, cited after Francis, 1986; Quirk, Greenbaum, & Leech, 1989). For structures with a single prepositional phrase modifier used here and in many earlier studies, linear precedence between candidate noun phrases seems to play a major role in the explanation of attraction effects during on-line processing (Fayol et al., 1994; also see Haskell & MacDonald, 2005).

However, it should be noted that the explanation in terms of a simple function of linear distance was contested by a number of studies that stressed the importance of the structural relation that holds between controller and attractor. Studies comparing different types and depths of embedding of attractors yielded evidence that the strength of the attraction effect also depends on the hierarchical distance between attractor and target (Vigliocco & Nicol, 1998; Hartsuiker, Méndez, & van Zee, 2001; Franck, Vigliocco, & Nicol, 2002), and on the type of interpolated structure (Franck, Lassi, Frauenfelder, & Rizzi, 2006; Franck, Soare, Frauenfelder, & Rizzi, 2010).

Apart from configurational factors, morphophonology appears to play a role during the on-line computation of agreement, in that the informativity or (un-)ambiguity of morphological number markings on head and local noun (and their respective determiners) can provide cues that influence the amount of agreement attraction, see for instance Hartsuiker et al. (2003); Badecker and Kuminiak (2007), or Franck et al. (2002). Other evidence points to the fact that conceptual (or 'notional') information can influence the production of agreement as well. Different studies have reported effects of plausibility (Thornton & MacDonald, 2003), natural gender (Deutsch & Dank, 2009), animacy (Bock & Miller, 1991), and distributivity (e. g. Vigliocco, Butterworth, & Semenza, 1995; Eberhard, 1999; Hartsuiker & Barkhuysen, 2006) on the prevalence of agreement errors.

Finally, *plural markedness* has been a recurrent, characteristic effect in studies that elicit agreement errors in English: When the actual agreement controller is marked singular and a potential attractor with plural marking is present, attraction errors are significantly more likely than in the opposite case (compare examples 2 and 3). Bock & Eberhard (1993) find almost twice as many agreement errors in an experiment using structures similar to example sentence 2, with a local noun marked for plural, compared to sentences like the one given in example 3 (note, however, that the overall number of responses with agreement error observed in this experiment is fairly small: 26 out of 768 observations).

(2) The inscription on the ancient pillars *is*/**are* weathered.

(3) The inscriptions on the ancient pillar **is*/*are* weathered.

Different authors have attributed plural number a 'marked' status over singular on phonetic, morphological, and conceptual or semantic grounds, see for instance Greenberg (1966/2005), Givón (1991), or Wurzel (1998).[1] A central claim of the explanation of the plural markedness effect by Eberhard (1997) is that number marking is based on a privative or unary feature specification, and that the featureless, 'unmarked', or *default* number is singular. A 'default' marking or singular bias has been suggested earlier by Hemforth and Konieczny (2003) as an explanation of their results from a written sentence completion study; also see Franck et al. (2004) for a similar argument based on acquisition data. Eberhard (1997) presents evidence for the psychological reality of the unary specification of number by a grammatical feature, which is in line with common linguistic accounts of number marking.

The current study

In order to test the influence of situational, external distraction on agreement generation, we conducted a study employing an adapted elicitation paradigm in which subjects had to produce sentences under noise, based on sentence preambles they were given as stimuli. In order to avoid noise masking during the perception of the stimuli, we opted for visual presentation of the preambles.

Based on the literature reviewed in the previous sections, we expected that speech-free noise will create a secondary task load that can interfere with language processing. The findings from the ISE literature further predicted that temporally structured (fluctuating) noise would lead to stronger disruptions than constant noise. We therefore used four different noise conditions: a silent baseline, steady-state noise, and two fluctuating noise signals with a different intensity of the fluctuation.

We are at present unaware of any previous studies that investigated the effects of external noise as a distractor on the production of agreement with the paradigm employed here. Therefore we had no reference point for a quantitative hypothesis about the effect of noise. Still, we expect the additional load generated by the presence of noise to impact the language production process. We assume that secondary task load effects are the result of a competition for processing time on procedures shared between primary and secondary task. Therefore we expect a slowing down of language processing, which should be reflected in higher error rates. In addition, on-line measures like utterance latency or speech rate should be affected, leading to longer latencies before speaking starts, and to slower articulation.

Sentence elicitation experiment

Method

For the experiment we used a sentence elicitation task with *rapid serial visual presentation* (RSVP) of sentence preambles (Potter, 1984). The visual presentation was necessary because part of the presentation would take place under noise, and the noise would partially mask auditorily presented stimuli, introducing an additional, severe source of error to the subject's answers. According to Potter (1984), the serial word-by-word presentation mode is more similar to listening than to conventional reading, given that looking ahead or back is not possible with this way of presentation. Results from studies, for instance by Potter and Lombardi (1998), indicate that subjects can accurately read and recall sentences shown with RSVP.

Participants

24 undergraduate and master students of the University of Oldenburg, aged between 20 and 28 years (M = 23.6, SD = 2.12), participated in the study; half of them were female. Subjects were recruited through an on-line bulletin board announcement, and were paid 7.50 Euro per hour for participation. All participants were naive with regard to the experimental manipulation. Before testing, each subject was screened for hearing acuity with pure-tone audiometry, carried out on an Interacoustics AC40 audiometer. All participants had normal hearing according to WHO (World Health Organization, n. d.) standard.

In order to increase statistical power, each participant was tested two times, with at least two weeks in between sessions. Only data from participants that took part in both experimental sessions were used for the analysis.

Material

20 German test sentences were constructed after the pattern used by Bock and Miller (1991). They consisted of a copula construction with an embedded prepositional phrase (PP) modifying the head of the subject noun phrase (also see Schriefers & van Kampen, 1993; Hemforth & Konieczny, 2003; Hartsuiker & Barkhuysen, 2006). Number marking on head and local NPs was manipulated to yield four different versions per sentence. For that, we combined the two factors *match* (head and local noun match/mismatch in number) and *head noun number* (singular/plural). The (intended) structure of the sentences was [$_S$[$_{NP}$ D N[$_{PP}$ P [$_{NP}$ D ADJ N]]] V$_{AUX}$ ADJ], see example 4.

(4) Die Inschrift-en$_{HEAD}$ auf der antik-en Säule$_{LOCAL}$ *sind* verwitter-t.

the inscription-PL on the ancient-SG pillar;SG are weather-PTCP

'The inscriptions on the ancient pillar are weathered.'

(5)	*	Die	Inschrift$_{HEAD}$	auf	den	antik-en	Säule-n$_{LOCAL}$	*sind*	verwitter-t.
		the	inscription;SG	on	the	ancient-PL	pillar-PL	are	weather-PTCP

'The inscription on the ancient pillars are weathered.'

In order to generate sentences, two nouns and an adjective were combined to form plausible propositions. In some cases where appropriate, German translations of material from earlier studies were used. Nouns and adjectives used in the sentences were controlled for word length in syllables, as well as for frequency class according to the *Leipziger Wortschatz* database (Biemann, Bordag, Heyer, Quasthoff, & Wolff, 2004), so that lexical elements within an individual sentences would not stand out on either of the two measures.

All head nouns used in the material had feminine gender. This was to avoid any additional cue from the determiner, which is usually marked for case, gender and number in German. The paradigm for the definite determiner shows syncretism between singular and plural for feminine gender in nominative and accusative case, hence the determiner would invariably have the form "die". For the embedded attractor phrase we used prepositions which select for either accusative or dative case and combined them with nouns of masculine, feminine or neutral gender. Also, a second adjective was added to the prepositional phrase, in order to increase the distance between subject head noun and verb. With the chosen combinations of preposition and noun, the determiner of the embedded noun was always unambiguously marked for singular or plural,[2] and therefore the form of the adjective did not change, which would be the case if combined with an ambiguously marked definite determiner. Additionally, some of the resulting noun-combinations allow a distributive reading, which has been shown to increase error rates (Vigliocco, Butterworth, & Garrett, 1996).[3]

To pre-test the items, we carried out a web-based questionnaire study with 90 participants (all native speakers of German, 23 male, 67 female; aged between 19 and 59, M = 24.6, SD = 5.21). The questionnaire asked for a plausibility rating about each sentence, given a scale ranging from 1 to 4. Based on the rating results, we removed items where the difference in plausibility between the four different head/local NP number combinations reached significance, and items where the overall standard deviation was above 0.9. This left us with 20 experimental items, with four sentence versions per item (see appendix A for a list of items). For the experiment, the four number marking conditions (singular match [sg-sg], plural match [pl-pl], singular mismatch [sg-pl], plural mismatch [pl-sg]) were counterbalanced across four noise condition blocks, so that one combination of nouns never appeared more than once within one block. Four experimental lists were created with different distributions of material across blocks.

The experimental lists were pseudo-randomly interspersed with filler items taken from the *OLACS* corpus of stimulus sentences (Uslar et al., 2010). Care was taken to limit the number of succeeding experimental items to three in all cases. The filler material consisted of 160 sentences, 80 relative clauses and 80 simple transitive clauses (see examples 6 and 7).

(6)	Der	freche	Kasper	tadelt	den	stolzen	Clown.
	the;NOM	cheeky	buffoon	reprimands	the;ACC	proud	clown.

'The cheeky buffoon is reprimanding the proud clown.'

(7)	Der	Soldat,	der	die	Köchinnen	tadelt,	schwitzt.
	the;NOM	soldier,	who	the;ACC	chefs;F	reprimands,	sweats.

'The soldier, who is reprimanding the female chefs, is sweating.'

Four different sentence types per filler structure were used: both SVO and OVS sentences with and without plural subject, as well as subject and object relative clauses with and without plural RC subject.

Each sentence in the experiment, including fillers, was 8 words long, and always the second to last word was replaced by an underscore that indicated the gap participants had to fill. Across critical and filler sentences, subjects were required to complete gaps with either an adjective or a full verb in the filler items, or a copula (auxiliary) verb for the critical items. In general, if a subject noticed the number manipulation in the material, what he or she experienced was variation of the number of different NPs across three different structures.[4]

Procedure

After briefing and screening, subjects were seated in front of a 19-inch LCD screen in a dimly lit sound attenuated booth. They were instructed to attend to a fixation cross on the screen and to a sentence containing a gap, which would be presented in a word-by-word fashion. Subjects were asked to read the sentence, wait for an answer prompt after the complete presentation and then speak aloud the sentence they had read, including their completion of the missing word.

(FIGURE 1 ABOUT HERE)

The exact timing of an individual trial is shown in figure 1. After a 1000 ms fixation cross the sentence appeared in a word-by-word fashion, containing a slide with underscores marking the gap. The end of the sentence was indicated by a period after the last word. Following the final word a visual mask appeared for 500 ms, consisting of a row of hash signs (#). The mask was followed by the production prompt, a single question mark at the centre of the screen. The prompt lasted for 4500 ms, during which subjects had been instructed to respond. After an ITI of 1500 ms the next trial started.

In the noise conditions a pre-recorded distractor noise sound was presented. The distractor sound started after the sixth word of a sentence and lasted until the end of the trial, crucially including the time slot allotted to the participants' response. This way we made sure that there was little noise interference with reading and memorising the sentence preamble, while still a good portion of the sentence planning and all of the actual articulation of the subjects' answers took place in noise.

Audiovisual stimulus presentation and sound recording was carried out with E-Prime Pro 2 software (Psychology Software Tools, Inc.). For distractor sound presentation, we used two Genelec 8030A active near-field speakers. The spoken responses of the participants were recorded with a low-latency sound adaptor (ECHO Audio Gina 3G) and an AKG C1000-S microphone, both attached to the presentation PC. The microphone we used has a cardioid recording characteristic and was directed at the speaker's mouth, so we could achieve a good signal-to-noise ratio of the recordings that would allow later scoring of the sentences for all noise conditions. Before the study, the experimental set-up was calibrated with a high-precision sound pressure level meter (Brüel & Kjær Investigator 2260). This way we determined fixed hard- and software settings with which we would reach an average (root mean square, RMS) intensity for the different noise files of 68 dB SPL at the approximate position of a participant's head.

Apart from silence, which served as baseline condition, we used three different noise signals as distractor stimuli. The signals are part of the *International Collegium of Rehabilitative Audiology* (ICRA) set of standardised noises for audiological research, simulating spectral and or temporal characteristics of human speech (Dreschler, Verschuure, Ludvigsen, & Westermann, 2001). We chose one unmodulated signal (ICRA 1), with long-term spectrum characteristics of speech ('pink' noise), a modulated signal simulating the prosodic contour of one speaker at small distance (ICRA 4), and a signal simulating six talkers speaking at the same time (ICRA 7). The ICRA 4 signal was used because experiments on language perception have found large negative effects on intelligibility with speech masked by this type of fluctuating sound (Wagener, Brand, & Kollmeier, 2006). We included the ICRA 7 'babble' sound as well, because it is conceptually closest to a "cocktail party" setting (Bronkhorst, 2000). For the actual presentation of the different noise sounds we used one 20-second snippets from each of the much longer original sound files. The noise conditions were blocked in order to reduce surprise effects. Block order was counterbalanced across participants and each participant received a different block order on each of the two experimental sessions.

Scoring

Only answers that conformed to the intended structure (see example 4) were analysed. The subjects' answers were scored manually based on the recordings, registering the produced number of both head and local noun as well as the number marked on the verb. An answer was scored as containing an *agreement error* if the respective number markings on subject head noun and verb did not match. Ungrammatical structures, sentences that did not conform to the intended structure, or unintelligible answers were scored as *other*. Instances where a subject had produced a different combination of head and local noun number than the one presented were scored as *repetition errors*, since we could not exclude the possibility that participants had misread or had incorrectly remembered the visually presented sentences. Further analyses of agreement errors were performed based on a data set from which *other* and *repetition errors* had been removed; analyses of repetition errors are based on a data set excluding *other* errors. Table 1 provides an overview of the different kinds of errors produced by participants.

(TABLE 1 ABOUT HERE)

In addition to the error counts we also collected data about the time course of the subjects' articulation. Each response had been recorded separately during the experiment, with the recording onset synchronised to the answer prompt. We gauged the utterance latency by manually measuring from the start of the recording to the onset of a subject's utterance, using the PRAAT software (Boersma & Weenink, 2010). Cases where a subject had begun speaking before the recording started were excluded. In order to measure speech rate, we also identified the onset of the *critical verb* in a subject's answer, and calculated the duration of the initial portion of each utterance. The length of this fragment should reflect the time course of the relevant planning processes up to the point where the articulation of the verb starts. Detailed orthographic transcriptions of the answers were made and served as the data base to count the number of syllables produced in each utterance up to the verb, in order to establish a speech rate measure in syllables per second.

Design and Analysis

Data from the two sessions per participant were pooled. The data were analysed using mixed effects models (Baayen, Davidson, & Bates, 2008; Baayen, 2008; Jaeger, 2008). To analyse the different error counts we computed generalised (logistic) mixed effects models with the respective error as outcome variable, and noise, match and head noun phrase number as predictors. Linear mixed effects models were computed for the continuous utterance latency and speech rate data. In both cases, subjects and stimulus sentences were treated as random effects by adding random intercepts to the models. To initially test the results from the *baseline* silence condition for a mismatch effect, a simple model was specified, based on previous results in the literature, with match condition and head noun number as the only predictors. The same model was then fitted to the entire data set from the experiment, first marginalising over noise conditions. Using the *anova()* function in the R software package (R Development Core Team, 2010; cf. also Baayen et al., 2008), an implementation of a log-likelihood ratio test comparing (Laplace) quasi log-likelihoods, we added noise as a factor and tested for possible interactions with noise by comparing it to the simpler model, as an omnibus test for the factor noise. P-values for linear mixed effects models were calculated with the *pvals.fnc()* function (Baayen et al., 2008).

Results

Figure 2 summarises the number of agreement errors as a function of the match and head NP number conditions in the four different noise conditions.

(FIGURE 2 ABOUT HERE)

When looking at the data from our *silent* condition only, we see that the number of errors is considerably higher in the *singular mismatch* condition compared to the other conditions. Statistically, the interaction between head noun number and match condition did not reach significance, while the effect of the factor match was significant (see model summary in table 2). However, applying the same model to the entire data set, marginalising over noise condition, yields a highly significant interaction effect between the *match* factor and head noun number (see table 3).

(TABLE 2 ABOUT HERE)

(TABLE 3 ABOUT HERE)

As described in the previous section, we fitted different models with and without noise condition as a predictor, in order to assess the influence of noise on agreement error incidence. As a result of the model comparison, we see that adding noise condition as a predictor significantly improved the model fit ($\chi^2(9) = 18.271$, $p<.05$). We see this as an indication that noise has an effect on the distribution of errors.

Crucially, two interactions of error likelihood with noise reached significance in the final model: When either of the two *fluctuating* noises was present, the number of errors increased in sentences with either of the two nouns marked for plural. Constant noise (ICRA 1) appears to not affect the error distribution. 3-way interactions between noise, match and subject noun number did not improve model fit and were removed from the specification. Table 4 summarises the final model including noise as a factor.

(TABLE 4 ABOUT HERE)

In our analysis of *repetition errors*, we did not find statistical evidence that noise would significantly improve model fit ($\chi^2(3) = 0.912$, p = .823). The baseline model shows a significant effect of head noun number condition, as well as a marginally significant effect of match condition, indicating that conditions which contain a plural marking are more prone to repetition errors than the singular match condition (see table 5 for a model summary). The distribution of *other* responses in our data did not appear to be influenced by noise or the different number markings in the sentence preamble.

(TABLE 5 ABOUT HERE)

The timing of subjects' utterances appears to be influenced by noise. Table 6 contains mean results of the measurements of utterance latency and speech rate by condition. Our model comparison results indicate that the addition of noise as a factor significantly improves model fit for both utterance latency ($\chi2(3) = 42.69$, $p<.001$) and speech rate ($\chi^2(3) = 29.687$, $p<.001$). Tables 7 and 8 summarise the final models. No interaction between the match and head noun number conditions and any of the noise conditions reached significance. Simple effects for each of the three noise conditions indicate that subjects show a *lower* latency and at the same time a *lower* speech rate when speaking under noise.

(TABLE 6 ABOUT HERE)

(TABLE 7 ABOUT HERE)

(TABLE 8 ABOUT HERE)

Discussion

Our analysis of agreement attraction effects for the entire data set shows a clear increase in the number of errors made in the singular mismatch condition (sg-pl), including a significant interaction effect between match condition and head noun number condition. This characteristic interaction, which is expected based on earlier studies, fails to reach significance in the silence baseline condition, however. At present, we have no clear explanation for this, but statistical power issues might play a role. The complete data set shows the characteristic asymmetry of errors occurring more frequently after a plural attractor noun, which replicates the plural markedness effect in attraction (Bock & Miller, 1991; Eberhard, 1997; Hartsuiker et al., 2003).

Crucially, the omnibus test for an effect of noise on the pattern of agreement errors indicates that subjects were influenced by the noise manipulation. This result speaks in favour of our first hypothesis that external noise can influence the sentence production process by creating a secondary task load. The results of our on-line measurements of speaking rate also warrant this interpretation: under noise, speakers articulate more slowly than in silence. The same effect has been observed before by Postma and Noordanus (1996), and provides strong evidence that speech-free distractor noise can cause interference that results in a slowing down of processing.

Our observations about the influence of noise on utterance latency seem a little counter-intuitive, indicating that subjects started their utterances *faster* with background noise present. A similar observation has been made in a study by Hanke, Hamann, and Ruigendijk (2012). At present, we can only speculate about the reasons for this decrease in reaction time under noise, as different causes might be play a role for this effect: for instance fluctuations in the arousal level of subjects, or strategies to alleviate the burden on memory imposed by the task.

An interesting outcome from the analysis of agreement error counts is the fact that only fluctuating noise conditions entered significant interactions with other factors, while we found no indication for constant noise to have a measurable effect on the distribution of agreement errors. This finding might be an indication that the different kinds of noise we used in the experiment differ in the severity of disturbance they cause during language production. Such an outcome would actually be expected based on results from the literature on the irrelevant sound effect (e. g. Klatte et al., 1995), where it has been suggested that detrimental effects on performance result from interference by changing-state noise, but not as strongly from interferences by constant noise. As the coefficients from the final model with noise as a factor indicate (see table 4), the distribution of errors across the MATCH by NUMBER conditions changes. Error rates increased significantly under fluctuating noise, whenever the head noun was plural.

The plural effect we observe means that under fluctuating noise subjects actually produced more verbs wrongly marked for singular after a plural head. This is surprising especially for the plural match (pl-pl) condition, where the plural markings on both the head noun and the attractor do not provide conflicting cues for assigning the correct number to the verb representation. If a plural local noun is assumed to exert an attraction effect, a spurious number assignment in the plural match condition, based on the local noun's number marking, should result in (coincidentally) *correct* number marking.

However, the plural effect we observe might be the result of an increase in 'default' singular number markings – an explanation that has been proposed in earlier studies, for instance by Hemforth and Konieczny (2003) or Franck et al. (2004). This argument gains indirect support by our analysis of the distribution of *repetition errors* in our data, that is the number of errors where subjects reproduced the sentence preamble with one or more incorrect number markings on the noun. While we did not find an indication that repetition errors were contingent on the noise condition, there is a statistically significant tendency for this kind of error to appear more often in conditions where more plural markings have to be recognised and/or remembered.

With respect to our agreement error data, how could (fluctuating) noise lead to a stronger reliance of subjects on the default or unmarked form? A first step towards an explanation might be provided by looking at models for the irrelevant sound effect (ISE). In particular, the *changing state hypothesis* by Jones and Macken (1993) and Macken, Tremblay, Alford, and Jones (1999) emphasises 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.

As we noted in the introduction, the task of producing sentences with correct agreement cannot be straightforwardly compared to the serial verbal recall tasks usually employed to investigate irrelevant sound effects. However, maintenance of the order of elements might be a basic function that serves on-line sentence generation and the computation of agreement at some point during processing. Competition for processing time on a seriation mechanism should lead to a slowing down of concurrent language processing. If under (fluctuating) noise processing is slowed down, retrieval of the agreement controller noun phrase can in some cases take too long, and an empty or 'default' number specification is used as a cue for the retrieval of the verb form. Alternatively, the plural cue might be set too late during verb form retrieval, so that the default verb form will win by means of a head start in activation because of its higher frequency. It should be possible to integrate both accounts with different existing models for verb agreement, for instance the Marking and Morphing model by Bock, Eberhard, and Cutting (2004), or the working memory retrieval model outlined in Badecker and Kuminiak (2007; also cf. Lewis & Vasishth, 2005 and Wagers et al., 2009). Whatever the precise mechanism, the effect of fluctuating noise we observe is not an effect of attraction as such, but rather follows from the defaulting to singular verb forms, which are inappropriate after a plural head noun. To put it differently, we might say that under the "difficult" conditions no agreement relation at all is established between a verb and its controller.

Conclusion

For the current study we successfully influenced the grammatical encoding stage during language production with non-speech noise. According to our current knowledge, this is the first time that the effect of an external, situational non-speech distractor on the production of agreement was established. As suggested by results from research on the irrelevant sound effect, it might be particularly noise with broad-scale intensity fluctuations that exerted a detrimental effect on the production of agreement. Crucially, the disturbance by noise did not only lead to an increase in agreement errors, but the choice of a default form seems to indicate that correct agreement is not established at all in some cases.

Our present data does not allow for a straightforward decision between different processing models for agreement production, but the effect we observed might add to the empirical base on which different models are being constructed. Importantly, our finding that something as common or unspecific as background noise can influence grammatical encoding speaks against the assumption that the formulation stage operates in a highly automatic fashion with dedicated, or 'exclusive' processing resources at its disposal (Ferreira & Pashler, 2002; Hartsuiker & Barkhuysen, 2006; Garrod & Pickering, 2007). The view that processing under dual-task conditions leads to interference between basic processing steps or 'functions' that are shared between cognitive tasks presents exciting opportunities to further investigate the complex processing mechanisms involved in production (and comprehension) of language, in order to dissociate domain-general and domain-specific processing functions (cf. for instance Barrett & Kurzban, 2006).

Notes

1. It should be borne in mind, however, that languages differ in the way they code number, and morphological complexity does not necessarily reflect greater conceptual complexity or markedness. In particular, there may be more to the plural markedness effects than for instance the frequency of an individual form alone, as (language-specific) properties of the respective verbal paradigms and properties of lexical access come into play here (see for instance Kostić, 1991, also see Milin, Filipović Đurđević, & Moscoso Del Prado Martín, 2009; Ewijk & Avrutin, 2010; Baayen, Levelt, Schreuder, & Ernestus, 2007). For German, we assume singular to be the preferred or 'default' number on morphological grounds and based on consideration of the lexical frequency of those verb forms used by subjects in almost all cases in our experiment.
2. Incidentally, the determiners were in most cases also unambiguously marked for (non-nominative) case. As one of the anonymous reviewers rightly pointed out, these linguistic factors can influence the likelihood of agreement errors (Hartsuiker et al., 2003). The main focus of the present experiment lay on the influence of noise on agreement errors, and noise was a within-item factor in our design. Therefore we will leave the question after potential *interactions* between the effects of noise and different linguistic cues for agreement computation open for future research.
3. As with case marking on the determiner, the semantic or conceptual difference between sentences with and without distributive reading was not central to the research goal of this paper. All sentence preambles were used under all noise conditions, and we will leave the question about potential interactions between noise and distributivity open at present.
4. Upon debriefing, all subjects reported that they had not been aware of the experiment's goal to elicit attraction errors, even if they had noticed the different number markings.

A. Stimuli

1. Die Beschwerde/n von der/den engagierten Schülerin/nen *ist/sind* begründet.

   "The complaint/s by the dedicated student/s *is/are* justified."

2. Die Grafik/en in der/den farbigen Broschüre/n *ist/sind* deutlich.

   "The figure/s in the colorful brochure/s *is/are* clear."

3. Die Inschrift/en auf der/den antiken Säule/n *ist/sind* verwittert.

   "The inscription/s on the ancient pillar/s *is/are* weathered."

4. Die Visite/n auf der/den chirurgischen Station/en *ist/sind* beendet.

   "The round/s of the surgical ward/s *is/are* finished."

5. Die Affäre/n um den/die berühmten Spion/e *ist/sind* übertrieben.

   "The scandal/s about the famous spy/spies *is/are* exaggerated."

6. Die Abbildung/en auf der/den begehrten Tasse/n *ist/sind* hübsch.

   "The picture/s on the sought-after cup/s *is/are* pretty."

7. Die Zeitschrift/en mit dem/den peinlichen Fehler/n *ist/sind* unauffindbar.

   "The newspaper/s with the embarrassing mistake/s *is/are* untraceable."

8. Die Lücke/n in der/den merkwürdigen Theorie/n *ist/sind* unübersehbar.

   "The gap/s in the awkward theory/theories *is/are* highly visible."

9. Die Kooperation/en mit dem/den gewerblichen Händler/n *ist/sind* fruchtbar.

   "The cooperation/s with the commercial dealer/s *is/are* fruitful."

10. Die Tracht/en mit dem/den typischen Muster/n *ist/sind* begehrt.

    "The traditional costume/s with the typical pattern/s *is/are* sought-after."

11. Die Beschreibung/en in dem/den aktuellen Reiseführer/n *ist/sind* undeutlich.

    "The description/s in the up-to-date travel guide/s *is/are* unclear."

12. Die Anweisung/en auf dem/den internen Rundschreiben/n *ist/sind* fehlerhaft.

    "The direction/s in the internal newsletter/s *is/are* inaccurate."

13. Die Manipulation/en an der/den komplizierten Maschine/n *ist/sind* gefährlich.

    "The manipulation/s of the complex machine/s *is/are* dangerous."

14. Die Anzeige/n in der/den erneuerten Bahn/en *ist/sind* beleuchtet.

    "The display/s in the renovated train/s *is/are* lit."

15. Die Versammlung/en vor der/den geschlossenen Filiale/n *ist/sind* unerwartet.

    "The gathering/s in front of the closed store/s *is/are* unexpected."

16. Die Unterschrift/en auf der/den verbindlichen Erklärung/en *ist/sind* gefälscht.

    "The signature/s on the binding declaration/s *is/are* forged."

17. Die Behandlung/en mit dem/den neuartigen Medikament/en *ist/sind* wirksam.

    "The treatment/s with the novel drug/s *is/are* effective."

18. Die Panne/n in der/den verarbeitenden Fabrik/en *ist/sind* katastrophal.

    "The accident/s in the processing factory/factories *is/are* catastrophic."

19. Die Veränderung/en an der/den gültigen Richtlinie/n *ist/sind* minimal.

    "The change/s to the effective regulation/s *is/are* minimal."

20. Die Forderung/en nach der/den umfangreichen Reform/en *ist/sind* unseriös.

    "The demand/s for the substantial reform/s *is/are* unsound."

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.

Baayen, R. H., Levelt, W. J. M., Schreuder, R., & Ernestus, M. (2007). Paradigmatic structure in speech production. *Proceedings from the Annual Meeting of the Chicago Linguistic Society, 43*(1), 1–29.

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. 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.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113*(3), 628–647.

Biemann, C., Bordag, S., Heyer, G., Quasthoff, U., & Wolff, C. (2004). Language-Independent Methods for Compiling Monolingual Lexical Data. *Lecture Notes in Computer Science, 2945,* 217–228.

Bock, J. K. (1982). Toward a Cognitive Psychology of Syntax: Information Processing Contributions to Sentence Formulation. *Psychological Review, 89*, 1–47.

Bock, K., & Levelt, W. J. M. (1994). Language Production: Grammatical Encoding. In M. A. Gernbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Bock, K., & Miller, C. A. (1991). Broken Agreement. *Cognitive Psychology, 23*(1), 45–93.

Bock, K., & Eberhard, K. M. (1993). Meaning, Sound and Syntax in English Number Agreement. *Language and Cognitive Processes, 8*, 57–99.

Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language, 51*, 251–278.

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

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica, 86*(1), 117–128.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*(1), 77–126.

Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language, 60*(1), 112–143.

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.

Eberhard, K. M. (1997). The Marked Effect of Number on Subject–Verb Agreement. *Journal of Memory and Language, 36,* 147–164.

Eberhard, K. M. (1999). The Accessibility of Conceptual Number to the Processes of Subject–Verb Agreement in English. *Journal of Memory and Language, 41*, 560–578.

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.

Ewijk, L. Van, & Avrutin, S. (2010). Article omission in Dutch children with SLI: A processing approach. *Entropy, 12,* 798–817.

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive Overload and Orthographic Errors: When Cognitive Overload Enhances Subject-Verb Agreement Errors. A Study in French Written Language. *The Quarterly Journal of Experimental Psychology, 47A*(2), 437–464.

Ferreira, V. S. (2007). How are speakers' linguistic choices affected by ambiguity? In A. S. Meyer, L. R. Wheeldon, A. Krott (Eds.), *Automaticity and control in language processing* (pp. 1–20). Hove: Psychology Press.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(6), 1187–1199.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Francis, W. N. (1986). Proximity Concord in English. *Journal of English Linguistics, 19,* 309–317.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition, 101*(1), 173–216.

Franck, J., Ohayon, S. Cronel, Chillier, L., Frauenfelder, U. H., Hamann, C., Rizzi, L., et al. (2004). Normal and pathological development of subject-verb agreement in speech production: a study on French children. *Journal of Neurolinguistics, 17,* 147–180.

Franck, J., Soare, G., Frauenfelder, U. H., & Rizzi, L. (2010). Object interference in subject–verb agreement: The role of intermediate traces of movement. *Journal of Memory and Language, 62,* 166–182.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes, 17*(4), 371–404.

Garrod, S., & Pickering, M. J. (2007). Automaticity of language production in monologue and dialogue. In A. S. Meyer, L. R. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 1–20). Hove: Psychology Press.

Givón, T. (1991). Markedness in grammar: distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language, 15*(2), 335–370.

Greenberg, J. H. (2005). *Language universals: with special reference to feature hierarchies*. Berlin: Mouton de Gruyter. (Original work published 1966)

Hale, W. G., & Buck, C. D. (1966). *A Latin Grammar*. Tuscaloosa, AL: University of Alabama Press. (Original work published 1903)

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

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes, 21*(1-3), 181–204.

Hartsuiker, R. J., Antón-Méndez, I., & van Zee, M. (2001). Object Attraction in Subject–Verb Agreement Construction. *Journal of Memory and Language, 45,* 546–572.

Hartsuiker, R. J., Schriefers, H. J., Bock, J. K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject–verb agreement. *Memory & Cognition, 31*(8), 1316–1326.

Haskell, T. R., & MacDonald, M. C. (2005). Constituent Structure and Linear Order in Language Production: Evidence From Subject–Verb Agreement. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 891–904.

Hemforth, B., & Konieczny, L. (2003). Proximity in agreement errors. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Boston, MA: Cognitive Science Society.

Hölscher, C., & Hemforth, B. (2000). Subject-verb agreement in German. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 279–310). Dordrecht: Kluwer.

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.

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 (ICASSP)* (Vol. 4). Phoenix, AZ: IEEE.

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.

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.

Kostić, A. (1991). Informational approach to the processing of inflected morphology: Standard data reconsidered. *Psychological Research, 53,* 62–70.

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., & 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.

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.

Meyer, A. S., Wheeldon, L. R., & Krott, A. (Eds.). (2007). *Automaticity and control in language processing*. Hove: Psychology Press.

Milin, P., Filipović Đurđević, D., & Moscoso Del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language, 60*(1), 50–64.

Moravcsik, E. (2006). *An Introduction to Syntax: Fundamentals of Syntactic Analysis*. London: Continuum.

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes, 17*(2), 163–184.

Pearlmutter, N. J., Garnsey, S. M., & Bock, J. K. (1999). Agreement Processes in Sentence Comprehension. *Journal of Memory and Language, 41,* 427–456.

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.

Potter, M. C. (1984). Rapid Serial Visual Presentation (RSVP): A Method for Studying Language Processing. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 91–118). Hillsdale, NJ: Erlbaum.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*(3), 265–282.

Quirk, R., Greenbaum, S., & Leech, G. N. (1989). *A comprehensive grammar of the English language* (7th ed.). London: Longman.

R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing [Computer software manual]*. Vienna, Austria. Retrieved from http://www.r-project.org

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.

Schriefers, H., & van Kampen, A. (1993). Syntaktische Prozesse bei der Sprachproduktion: Zur Numerus-Kongruenz zwischen Subjekt und Verb [Syntactic processes in sentence production: On number agreement between subject and verb]. *Sprache & Kognition, 12*(4), 205–216.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language, 48*(4), 740–759.

Uslar, V., Brand, T., Hanke, M., Carroll, R., Ruigendijk, E., Hamann, C., et al. (2010). Does sentence complexity interfere with intelligibility in noise? Evaluation of the Oldenburg Linguistically and Audiologically Controlled Sentence Test (OLACS). In *Proceedings of the 11th Annual Conference of the International Speech Communication Association INTERSPEECH 2010* (pp. 2482–2485). Makuhari, Japan: International Speech Communication Association.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. Journal of Memory and Language, 34, 186–215.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61,* 261–289.

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: is proximity concord syntactic or linear? *Cognition, 68*(1), B13–B29.

Wagener, K. C., Brand, T., & Kollmeier, B. (2006). The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology, 45*(1), 26–33.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61,* 206–237.

World Health Organization. (n. d.). *Prevention of blindness and deafness: Grades of hearing impairment*. Retrieved from http://www.who.int/pbd/deafness/hearing_impairment_grades/en/index.html

Wurzel, W. U. (1998). On Markedness. *Theoretical Linguistics, 24*(1), 53–71.

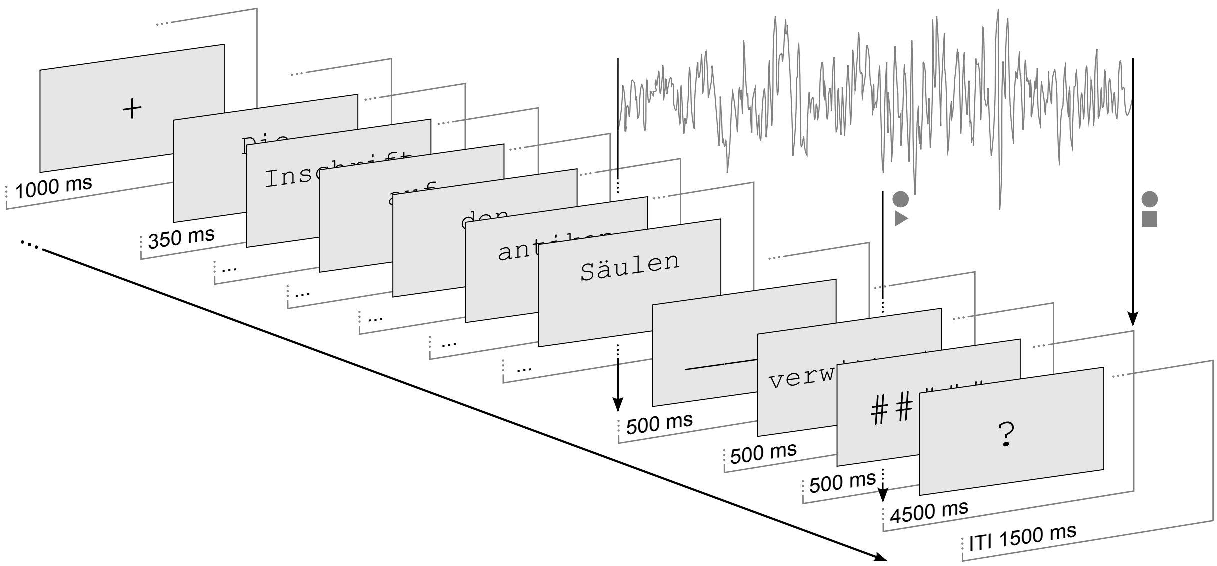

Figure 1: Trial scheme for RSVP stimulus presentation. Vertical arrows indicate the respective timing of noise presentation and answer recording.

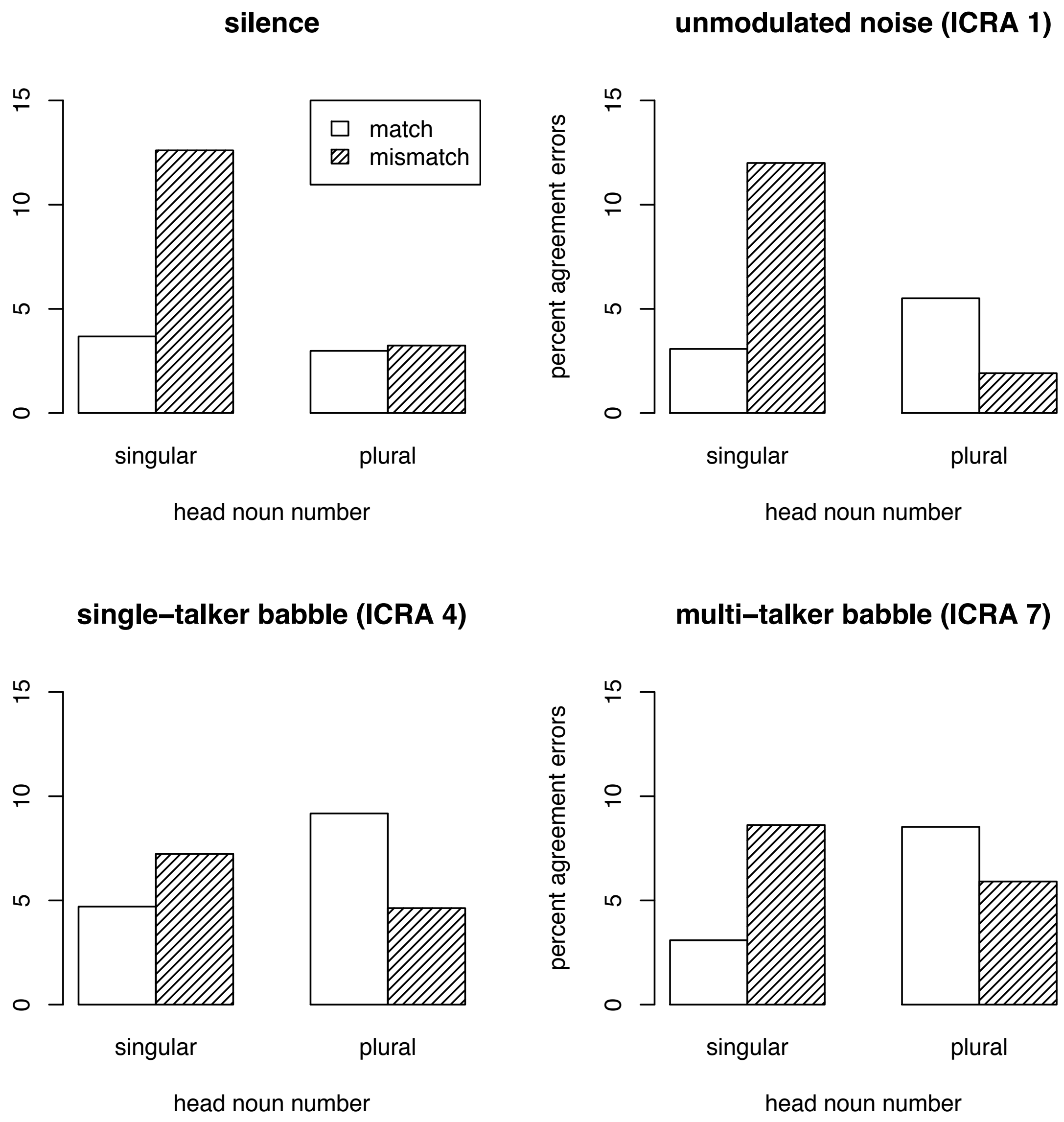

Figure 2: Error percentage under different noise conditions.

Table 1: Error counts by condition, total number of recorded answers $N = 3840$. For each combination of noise condition, head noun number and match condition, 240 answers were recorded. The percentages in figure 2 were calculated by dividing the number of agreement errors by the total number of answers per cell, excluding 'other' answers and repetition errors.

	Match condition	mismatch		match		
	Head noun no.	singular	plural	singular	plural	
Noise condition	Error type					Row sums
Silence	Agreement	24	5	6	3	38
	Repetition	43	45	23	64	175
	Other	11	7	11	8	37
ICRA 1	Agreement	23	3	3	9	38
	Repetition	47	53	25	42	167
	Other	10	8	4	7	29
ICRA 4	Agreement	16	7	6	16	45
	Repetition	37	43	31	50	161
	Other	10	11	10	7	38
ICRA 7	Agreement	16	9	8	14	47
	Repetition	36	52	27	52	167
	Other	14	6	6	13	39
Col. sums	Agreement	79	24	23	42	
	Repetition	163	193	106	208	
	Other	45	32	31	35	

Table 2: GLMM coefficients for the agreement error data from the silent baseline condition, predictors are treatment-coded (N = 748; log-likelihood = –133.7).

	Coef β	SE(β)	z	p
(Intercept)	–3.93	0.51	–7.8	**<.001**
Head number:PL	–0.52	0.81	–0.6	.519
Match:MISMATCH	1.71	0.54	3.1	**<.01**
Hd no.:PL × Match:MISMATCH	–1.33	1.00	–1.3	.181

Table 3: GLMM coefficients for the agreement error data for the full data set (predictors are treatment-coded; $N = 3026$; log-likelihood = –576.2).

	Coef β	SE(β)	**z**	**p**
(Intercept)	–4.16	0.36	–11.7	**<.001**
Head number:PL	0.87	0.40	2.2	**<.05**
Match:MISMATCH	1.62	0.38	4.2	**<.001**
Hd no.:PL × Match:MISMATCH	–2.36	0.55	–4.3	**<.001**

Table 4: Coefficients for GLM model with noise as factor, predictors are treatment-coded (N = 3703; log-likelihood = –761).

	Coef β	SE(β)	**z**	**p**
(Intercept)	–4.28	0.50	–8.5	**<.001**
Head number:PL	–0.03	0.59	–0.1	.952
Match:MISMATCH	2.08	0.53	3.9	**<.001**
Noise:ICRA1	–0.08	0.56	–0.1	.885
Noise:ICRA4	0.25	0.53	0.5	.641
Noise:ICRA7	0.22	0.53	0.4	.673
Hd no.:PL × Match:MISM.	–2.24	0.57	–3.9	**<.001**
Hd no.:PL × Noise:ICRA1	0.50	0.60	0.8	.412
Hd no.:PL × Noise:ICRA4	1.29	0.57	2.3	**<.05**
Hd no.:PL × Noise:ICRA7	1.21	0.57	2.1	**<.05**
Match:MISM. × Noise:ICRA1	–0.18	0.59	–0.3	.766
Match:MISM. × Noise:ICRA4	–0.89	0.56	–1.6	.111
Match:MISM. × Noise:ICRA7	–0.77	0.56	–1.4	.169

Table 5: GLMM coefficients for the repetition error data for the full data set (predictors are treatment-coded; *N* = 3026; log-likelihood = –576.2).

	Coef β	SE(β)	**z**	**p**
(Intercept)	–2.44	0.28	–8.7	**<.001**
Head number:PL	0.87	0.32	2.7	**<.01**
Match:MISMATCH	0.58	0.32	1.8	.072
Hd no.:PL × Match:MISMATCH	–0.67	0.45	–1.5	.140

Table 6: Mean utterance latency (in seconds) and mean speech rate (syllables per second) by condition. Total number of analysed recordings $N = 2922$.

	Match condition	mismatch		match		
	Head noun no.	singular	plural	singular	plural	
Noise condition	Measure					Row avgs
Silence	Latency	.5956	.5958	.5842	.5629	.5887
	Speech rate	5.898	6.041	5.932	6.026	5.977
ICRA 1	Latency	.5511	.5176	.5604	.5327	.5425
	Speech rate	5.778	5.817	5.876	5.824	5.833
ICRA 4	Latency	.5300	.5019	.5301	.5222	.5209
	Speech rate	5.700	5.716	5.866	5.838	5.781
ICRA 7	Latency	.5487	.5475	.5424	.5345	.5429
	Speech rate	5.840	5.857	5.852	5.904	5.862
Col. avgs	Latency	.5565	.5412	.5548	.5375	
	Speech rate	5.804	5.859	5.882	5.894	

Table 7: LME model summary for utterance latency; N = 2887; log-likelihood = 675.9

	Coef β	SE(β)	t	**p**
(Intercept)	0.590	0.02	24.74	**<.001**
Head number:PL	–0.016	0.01	–1.19	.237
Match:MISMATCH	0.003	0.01	0.19	.849
Noise:ICRA1	–0.045	0.01	–4.70	**<.001**
Noise:ICRA4	–0.061	0.01	–6.25	**<.001**
Noise:ICRA7	–0.042	0.01	–4.26	**<.001**
Hd no.:PL × Match:MISM.	–0.003	0.02	–0.13	.886

Table 8: LME model summary for speech rate; $N = 2920$; log-likelihood = –3148

	Coef β	SE(β)	*t*	**p**
(Intercept)	5.985	0.13	45.58	**<.001**
Head number:PL	0.054	0.12	0.45	.605
Match:MISMATCH	–0.079	0.12	–0.66	.456
Noise:ICRA1	–0.146	0.04	–4.13	**<.001**
Noise:ICRA4	–0.183	0.04	–5.16	**<.001**
Noise:ICRA7	–0.105	0.04	–2.96	**<.01**
Hd no.:PL × Match:MISM.	0.051	0.17	0.30	.746