Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Recognition of Minimal Pairs in (un)predictive Sentence Contexts in two Types of Noise

Permalink
https://escholarship.org/uc/item/70z995v4

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
van Os, Marjolein
Kray, Jutta
Demberg, Vera

Publication Date
2021

Peer reviewed

Recognition of Minimal Pairs in (un)predictive Sentence Contexts in two Types of Noise

Marjolein van Os (vanos@coli.uni-saarland.de)
Department of Language Science and Technology, Saarland University
Campus C7.2, 66123 Saarbrücken, Germany

Jutta Kray (j.kray@mx.uni-saarland.de)
Department of Psychology, Saarland University
Campus A1.3, 66123 Saarbrücken, Germany

Vera Demberg (vera@coli.uni-saarland.de)
Department of Language Science and Technology;
Department of Computer Science, Saarland University
Campus C7.2, 66123 Saarbrücken, Germany

Abstract

Top-down predictive processes and bottom-up auditory processes interact in speech comprehension. In background noise, the acoustic signal is degraded. This study investigated the interaction of these processes in a word recognition paradigm using high and low predictability sentences in two types of background noise and using phonetically controlled contrasts. Previous studies have reported false hearing, but have not provided insight into what phonetic features are most prone to false hearing. We here systematically explore this issue and find that plosives lead to increased false hearing compared to vowels. Furthermore, this study on German for the first time replicates the overall false hearing effect in young adults for a language other than English.

Keywords: predictability; background noise; listening comprehension; false hearing

Introduction

When listening to speech, there are at least two sources of information available that help us decode the speaker's message: there is the sensory information in the form of the acoustic speech signal, and there is also contextual information that can help guide predictions (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990).

The aim of the present study was to examine whether acoustic misunderstandings that can arise from too strong reliance on contextual information are influenced by the type of noise (babble vs. white noise) and the sound characteristics of words (phonetic minimal pairs, that is, tense vs. lax vowels and plosives which differ in the place of articulation). Our study will reveal new insights regarding misunderstandings due to predictive context and the phenomenon of false hearing, which is characterized by high confidence in the correctness of a response, although answering incorrectly.

Related Work

Predictability and Noise Studies have shown that predictions can facilitate auditory processes in difficult listening situations like background noise (Dubno, Ahlstrom, & Horwitz, 2000; Hutchinson, 1989; Pichora-Fuller, Schneider, & Daneman, 1995; Sommers & Danielson, 1999). By making use of the information provided by the sentence context combined with the information that can be glimpsed in the acoustic signal, the adverse effects of difficult listening situation can even be overcome (Wingfield, Tun, & McCoy, 2005; Wingfield, Tun, & Rosen, 1995).

Relying on context can sometimes also lead to what is known as *false hearing*. In these cases, the listener mishears a word, but is convinced they identified it correctly. Studies manipulating the predictability of sentences found that relying on contextual information in incongruent conditions leads to false alarms of around 30% for younger adults, and up to 40% for older adults (Failes, Sommers, & Jacoby, 2020; Sommers, Morton, & Rogers, 2015). Other studies used a priming paradigm to manipulate predictability (Rogers, 2017; Rogers, Jacoby, & Sommers, 2012; Sommers et al., 2015) and found similar results. The aim of the present study is to replicate these basic findings on how misunderstandings and false hearing vary with sentence predictability and extend those to the German language, as the aforementioned studies were all conducted in English.

Type of Noise Background noise has a negative effect on speech comprehension through energetic masking. Both the speech signal and the competing noise have energy in the same frequency bands at the same time (Brungart, 2001). The acoustic cues that listeners need for sound identification are masked by the noise. If the background noise is competing speech, its acoustic cues can "attach" themselves to the target speech (Cooke, 2009). In this way, the type of noise, for example, white noise, babble noise, or competing speech from a single speaker, might have different effects on the target speech. Multi-speaker babble noise approximates the average long-term spectrum of the speech of an adult male speaker, whereas white noise has a flat spectral density with the same amplitude throughout the audible frequency range. As such, these two types of noise lead to contrastive interference effects as they obscure different parts of the speech signal. The present study used both babble noise and white noise to investigate the effects of the different levels of energetic masking. As babble noise has more energetic masking of a speech signal than white noise, we expect more misunderstanding and lower confidence in responses.

Various studies have compared speech intelligibility in babble noise and in white noise. Gordon-Salant (1985) tested 57 CV sequences embedded in multispeaker babble and compared the results of the listening experiment to similar studies with white noise maskers (e.g., Soli & Arabie, 1979). The results showed that while the interference effects of both types of noise are contrastive, the predominant acoustic cues used for perception of consonants are the same in babble noise and white noise. In a paradigm using both meaningful and nonsense words embedded in speech noise, babble noise, and white noise, Taitelbaum-Swead and Fostick (2016) found lower accuracy for white noise than for the two speech-related noises. However, other studies found worst performance for babble noise compared to steady-state noise (Garcia Lecumberri & Cooke, 2006; Simpson & Cooke, 2005), suggesting the effect of noise type depends also on task and population. However, no studies so far directly compared the effect of noise type on mishearings guided by the sentence context.

Sound contrast Besides different noise types, there is also evidence that the characteristics of speech sounds influence speech understanding differently. Investigating specifically voicing and place of articulation of plosives in audio-visual speech perception, Alm et al. (2009) found that voicing information cues are more robust in white noise and more susceptible to babble noise, and that the place of articulation cues is more susceptible to white noise than babble noise. Furthermore, plosives consist of a closure of some part of the vocal tract, followed by a short burst of energy. This burst can easily be masked by noise. Vowels generally have a longer, steadier signal, that can be easier to distinguish in background noise. Their energy primarily lies between 250 and 2000 Hz (first and second formant, Flanagan, 1955) and thus is lower than that of plosives, for which higher formants are also important for identification (Alwan, Jiang, & Chen, 2011; Edwards, 1981). Spectral frequency information has been found to be particularly important for identifying the place of articulation in plosives, which is the feature of interest in the current study (Edwards, 1981; Liberman, Delattre, Cooper, & Gerstman, 1954). In the present study, a distinction is made between plosives that differ in place of articulation, and vowels that are either tense or lax. Previous studies (Failes et al., 2020; Sommers et al., 2015) did not systematically vary specific aspects of the minimal pair of sound changes in their stimuli: they replaced the first or last phoneme, but did not control in how many or which phonetic aspects the phoneme is changed.

Study Goals and Predictions

The present study aims to investigate in more detail than previous work to what extent mishearing and false hearing are affected by different types of noise and how the noise affects different types of sounds, as little is known about how misunderstandings and confidence ratings are further modulated by noise type and the characteristic of sounds.

We hypothesize that in adverse listening conditions (e.g., added noise) participants rely more on the sentence context rather than the acoustic signal, and use the context to compensate for the increased processing costs of the sound, in line with prior findings on English.

Our second hypothesis is that there will be a difference between the two types of noise, specifically, that babble noise is a more difficult listening condition than white noise. This is in line with previous studies (Garcia Lecumberri & Cooke, 2006; Simpson & Cooke, 2005), and we assume to find the same overall effect for German.

Finally, we hypothesize that words with vowel contrasts are more easily identified correctly than words with plosive contrasts, due to the longer signal of vowels. In white noise, particularly, place of articulation cues in plosives are hard to identify (Alm et al., 2009), which should lead to fewer correctly identified words among plosives.

Method

Participants 48 native speakers of German were recruited using a crowd-sourcing platform (31M, mean age = 24 years) to participate in the experiment. None of the participants reported hearing difficulties.

Materials We constructed sentences based on German minimal pairs that had a contrast in the middle of the word. These contrasts were plosives differing in place of articulation or tense vs lax vowels, each accounting for approximately half of the stimuli. The sentence-final target words were predictable based on the preceding sentence context (mean cloze 0.72, high predictability condition, HP). Sentences for the low predictability condition (LP) were constructed by swapping the target word of the HP sentence with its partner from the minimal pair. This procedure let us investigate whether listeners could rely on small acoustic cues for word recognition, even in background noise, while keeping sentence contexts equal across conditions. Example stimuli with translations can be found in Table 1. We tested a total of 480 sentences, half of which were highly predictable and half of which were unpredictable.

Recordings were made of all high predictability sentences and were read by a female native speaker of German. Recordings of the low predictability sentences were constructed via cross-splicing using Praat (Boersma & Weenink, 2021, version 6.1.05) to ensure the intonation and stress patterns were identical across conditions and not indicative of the unpredictable items. Subsequently, all sentences were embedded in two types of background noise, white noise and a multi-speaker babble. In the babble noise, none of the speakers were intelligible to prevent informational masking (café noise, BBC Sound Effects Library, Crowds: Interior, Dinner-Dance, http://bbcsfx.acropolis.org.uk/), and the sample was chosen to minimize environmental noises. All items had a Signal to Noise ratio (SNo) of -5 dB, meaning the background noise was five dB louder than the target sound, based on the mean intensity of the target word. There was 300 ms of leading and trailing noise so that participants had a chance to get used to the noise before the speech started. We used the unmasked recordings as a control condition ("Quiet").

Procedure In the experiment, participants listened to the sentence and had to report the final word that they heard. The sentence minus the target word, was presented on the screen in written form to ensure there was contextual information that could be used by the participant even when the background noise made it difficult to understand the speech itself. Additionally, participants rated their confidence in having given the correct response on a scale of 1 (completely uncertain, guessed) to 4 (completely certain). We collected these confidence ratings as a control measure to investigate participants' awareness of possible mistakes they were making. The next trial started when the participant had filled in both questions and clicked on 'next' to proceed. Different experimental lists were constructed so that each participant saw a total of sixty items, half of which were HP items and half were LP items. Participants saw only one item out of a set of four, but all items occurred in each noise condition across the lists. The three noise conditions (white, babble, and quiet) were presented in blocks of twenty items with the quiet condition last, so that the goal of the experiment would not be immediately obvious to participants. The order of the white and babble noise conditions was counterbalanced across participants.

Table 1: Example stimuli.

1A Am Pool im Hotel gab es nur noch eine freie **Liege**.
HP *At the pool in the hotel there was only one free* ***lounger*** *left.*
1B Nach vier Jahren heiratete Paul seine große **Liebe**.
HP *After four years, Paul married his big* ***love****.*
1C Am Pool im Hotel gab es nur noch eine freie **Liebe**.
LP *At the pool in the hotel there was only one free* ***love*** *left.*
1D Nach vier Jahren heiratete Paul seine große **Liege**.
LP *After four years, Paul married his big* ***lounger****.*

Note. Highly predictable sentences (HP) were made based on minimal pairs (*Liebe* / *Liege*) in 1A and 1B), then sentence-final target words were swapped to make low predictability items (LP) with the sentence frames of 1A and 1B, resulting in 1C and 1D. English translations of the sentences have been given in *italics.*

Analysis Responses were coded on whether they matched the auditorily presented word (e.g., in example 1A in Table 1 "Liege" / "lounger", *target*), the similar sounding *distractor* (e.g., in 1A "Liebe" / "love"), or were a different word entirely (e.g., in 1A "Platz" / "space", *wrong*). To get a better idea of whether participants relied on the sentence context or on the speech signal, we coded the semantic fit of the incorrect responses (fitting or not fitting), as well as the phonetic distance between the incorrect responses and target and distractor items. We made phonetic transcriptions based on the Deutsches Aussprachewörterbuch (German Pronunciation Dictionary; Krech, Stock, Hirschfeld, & Anders, 2009) and calculated the weighted feature edit distance using the Python package *Panphon* (Mortensen et al., 2016). This distance was normalized by dividing it by the longest of the two compared words. The normalized distance fell between 0 and 1.

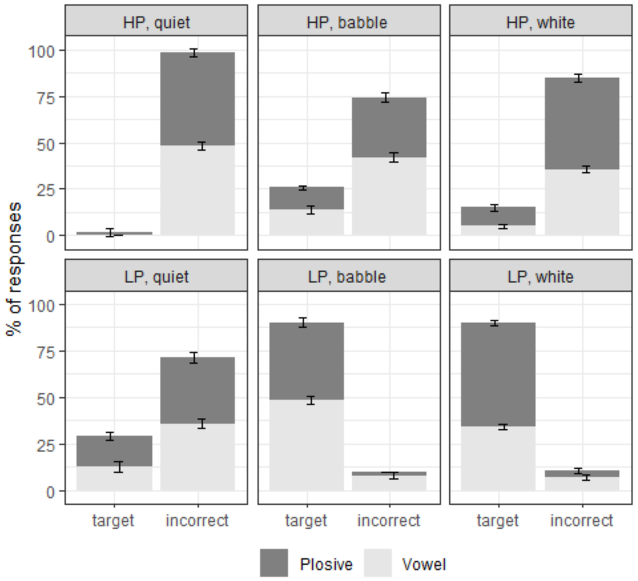

Figure 1: Percentage of Target and Incorrect (Wrong + Distractor) responses for both the High Predictability (HP) and Low Predictability (LP) condition, in quiet, babble, and white noise, split for plosives (P) and vowels (V).

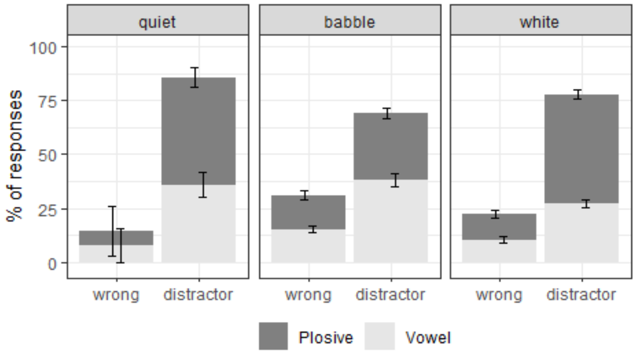

Figure 2: Percentage of Wrong and Distractor responses for the Low Predictability (LP) condition, in quiet, babble, and white noise, split for plosives (P) and vowels (V).

Results

We used general linear mixed models (GLMM), implemented in the lme4 package (Bates, Maechler, & Bolker, 2012) in R (R Development Core Team) to analyze our data. We analyzed the participants' binomial responses (0 = distractor/wrong, 1 = target) using a GLMM with a logistic linking function. To achieve convergence more easily, all models were run using the bobyqa optimizer and increased iterations to $2 \cdot 10^5$. The random structure of the models was reduced in cases of non-convergence. Model comparisons were made to guide model selection based on the Akaike Information Criterion (AIC), models with the lowest AIC are reported below.

Target vs. Incorrect Responses

To first replicate previous findings with our German materials, namely that the percentage of incorrect responses is higher for noisy as compared to quiet sentences, and whether this effect varies with predictability and noise type, we aggregated distractor and wrong response in our initial analyses. The corresponding data are displayed in Figure 1.

The model included fixed effects of Noise (categorical predictor with three levels, mapping Quiet to the intercept), Predictability (categorical predictor with two levels, mapping HP to the intercept), and Sound Contrast (categorical predictor with two levels, mapping Plosive to the intercept), and Trial No (continuous predictor, scaled). Trial No was added to control for learning effects over the course of an experimental block. Furthermore, the model included by-Participant and by-Item random intercepts, both with random slopes for Noise and Predictability. As expected, the model showed that both the Babble and White Noise conditions lead to higher incorrectly target responses than the Quiet condition ($\beta = -5.87$, $SE = 0.54$, $z = -10.87$, $p < .001$ for Babble and $\beta = -5.01$, $SE = 0.49$, $z = -10.15$, $p < .001$ for White Noise). A comparison of the two noise conditions shows that white noise led to more correct responses than babble noise ($\beta = 0.854$, $SE = 0.40$, $z = 2.15$, $p < .05$). Regarding the beneficial effect of predictability, we found that participants made more incorrect responses in the LP condition than in the HP condition ($\beta = -6.62$, $SE = 0.59$, $z = -11.21$, $p < .001$). We also found an effect of Sound Contrast where Vowel contrasts lead to more target responses than Plosive contrasts ($\beta = 0.82$, $SE = 0.23$, $z = 3.62$, $p < .001$).

Effect of Noise Type on the Type of Errors

Next, we took the subset of LP items, and tested whether the types of errors (distractor vs wrong) differ between the noise and sound contrast conditions. In these items, the context supports the distractor item, which is the minimal pair from the auditorily presented target word. As such, there is also some acoustic information supporting the distractor, and only careful listening would lead to reporting the target word. The model included fixed effects of Noise, Trial No, and Sound Contrast (all coded and scaled as before), as well as random intercepts for Participant and Item.

We found a significant effect of noise, indicating more wrong responses in Babble compared to Quiet ($\beta = -1.22$, $SE = 0.34$, $z = -3.56$, $p < .001$, White Noise did not differ significantly from Quiet, $p = .13$), as well as a significant difference between White Noise and Babble: White Noise leads to more distractor responses than wrong responses ($\beta = 0.70$, $SE = 0.25$, $z = 2.89$, $p < .01$). These effects are presented in Figure 2.

Controlling the Semantic Fit and Phonetic Distance

We coded the semantic fit and phonetic distance to the target of the wrong responses, to see whether participants for these responses relied more on the acoustic signal (low distance) or on the provided context (wrong response fits semantically). Figure 3 presents the normalized phonetic distance and semantic fit for the wrong responses in each of the three noise conditions. Lower normalized phonetic distance scores mean that the participant's response sounded more similar to the target word. Responses with a distance score of 1 were empty responses. We see that in all noise conditions the majority of the wrong responses did not fit the sentence semantically, suggesting participants did try to rely on the acoustic signal rather than the provided context. Especially in white noise (but also suggested by the right-side tail in babble noise) there is higher peak at larger phonetic distances for the responses that semantically fit the sentence, suggesting a trade-off between acoustic fit and semantic fit. When not responding with the target or distractor, participants made their response based on what they heard at a cost of fitting the semantic context. This is also the case in the quiet condition, where virtually none of the responses fit the context semantically but show a small phonetic distance to the target word, suggesting that participants chose to rely on the acoustic signal rather than the presented context.

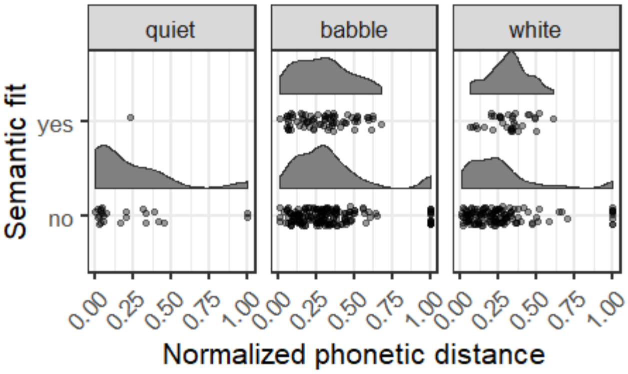

Figure 3: Wrong responses that semantically fit or did not fit the sentence, plotted with the normalized phonetic distance, in each of the three noise conditions.

Confidence Ratings

We had collected the participants' confidence ratings for each trial, and took two subsets of the data for the analyses: one with the wrong responses (N = 445), and one with the distractor responses (N = 760). We transformed the participants' confidence responses to a binary variable of low confidence (confidence ratings 1 and 2) and high confidence (confidence ratings 3 and 4). For the subset of distractor items, the model included fixed effects of Predictability, Noise, and the interaction, Trial No, and Sound Contrast (all coded and scaled as before). There were random intercepts for Participant and Item, with a random slope of Noise for Participant.

We find a significant effect of predictability, with higher confidence for distractor responses in LP sentences compared to HP sentences ($\beta = 2.64$, $SE = 0.96$, $z = 2.75$, $p < .01$). There was a significant effect of Noise in the case of Babble noise compared to Quiet, with lower confidence ratings in the Babble condition ($\beta = -1.57$, $SE = 0.63$, $z = -2.50$, $p < .01$). The effect did not reach significance between Quiet and White noise ($p = .15$). Comparing Babble and White Noise, the effect approached significance, suggesting participants were more confident in White Noise ($\beta = 0.61$, $SE = 0.31$, $z = 1.95$, $p = .05$). The model revealed a significant effect of Trial No, indicating that participants got less confident as the experiment went on ($\beta = -0.38$, $SE = 0.12$, $z = -3.21$, $p < .01$). Additionally, there was a significant effect of Sound Contrast ($\beta = -0.57$, $SE = 0.27$, $z = -2.09$, $p < .05$), suggesting that participants were less confident about their answers on items that had a vowel contrast, rather than those with a plosive contrast.

The model for the subset of wrong responses included the same fixed effects as the previous model, but now with additional Semantic Fit, Normalized Distance and the interaction. The random effects structure only included a random intercept for Noise, all other random effects led to singularity issues. The model revealed a significant effect of Noise: Participants were less confident in their responses in both types of noise ($\beta = -2.55$, $SE = 0.61$, $z = -4.19$, $p < .001$ for Babble and $\beta = -2.99$, $SE = 0.63$, $z = -4.74$, $p < .001$ for White Noise). There was no significant difference between the two noise conditions ($p = 0.16$).

Discussion & Conclusion

The present study investigated acoustic misunderstandings that can arise when listeners rely too strongly on contextual information. It aimed to extend previous studies on mishearing by investigating whether those effects are modulated by different types of noise and sound characteristics.

Effect of Predictability and Noise

Performance was best in HP sentences without noise as both the acoustic information and the contextual information can be easily processed and point to the same lexical candidate. Hence, in line with previous findings that contextual information supports word recognition (Dubno et al., 2000; Hutchinson, 1989; Pichora-Fuller et al., 1995; Sommers & Danielson, 1999), we showed that the effects can be replicated with German sentences. Our findings support the view of a trade-off between focusing on acoustic and semantic information; while participants strongly rely on the acoustic signal in favorable listening conditions, they turn more to the sentence context in background noise. This is the case in both babble and white noise, where listeners relied more on the sentence context, leading to incorrect responses in the LP condition. We thus replicated the effect of mishearing in younger adults in German, which has previously only been reported for English (Failes et al., 2020; Sommers et al., 2015).

The larger number of incorrect responses in babble noise suggests that this type of noise is the more difficult condition due to its large overlap with the frequencies of speech, obscuring the signal more than white noise, which is spread out over more frequencies. These findings suggest that in white noise, more of the speech signal is preserved that can guide participants' predictions than in babble noise. We find this effect both overall as in the low predictability subset (for wrong responses). It has been found by previous studies as well (Garcia Lecumberri & Cooke, 2006; Simpson & Cooke, 2005; but see Taitelbaum-Swead & Fostick, 2016 for opposite results). The wrong responses in the babble condition cannot have been caused by competing speech in the noise: due to the high number of speakers, specific speech streams were unintelligible. As such, the larger proportion of wrong responses compared to distractor responses must be due to interference of the noise itself.

Finding a larger amount of target responses for items with vowel contrasts suggests that words containing a vowel contrast were easier to identify correctly because they have a longer and steadier acoustic signal than plosives, which are characterized by their short burst. Thus, confirming our hypotheses, vowels are easier to identify than plosives.

Confidence Ratings

False hearing is defined by high confidence in incorrect responses. We find that different sound types lead to different degrees of false hearing, showing that false hearing depends not only on age or context, but also on acoustic characteristics of the message. In particular, plosives are more prone to false hearing than vowels: when the mishearing was based on a change in place of articulation in the plosive, participants reported higher confidence in their incorrect responses than when the misheard sound was a vowel.

A wrong response led to low confidence scores in general: 75%, while for distractor responses this was 22%. This suggests that participants were able to subjectively rate their confidence, as the distractor, which was predicted by the LP sentence context and mostly supported by the audio signal as well, would not lead to much doubt in the participants, whereas there was less support from both types of information for the wrong response, leading to lower confidence scores. The higher confidence for the distractor responses suggests false hearing, as these words were incorrect. This is also supported by the finding that the distractor response got higher confidence ratings in low predictability sentences compared to high predictability sentences. In low predictability sentences, the distractor fit the sentence semantically, but not in the high predictability sentences. Participants became less confident of their responses as the experiment went on. This might be because they started to notice the manipulation, and realized they could not rely on the sentence context in half of the trials (the low predictability items).

We saw before that words with a vowel contrast were more often identified correctly. If participants are less confident of their responses, they may focus their attention more on those targets, which in turn leads to better performance. However, this would have to be the case for both sound types, as only during presentation they could be aware of the sound contrast.

Language & Noise Type

Most studies on mishearing and false hearing have been done on English (Failes, et al., 2020; Rogers, 2017; Rogers, et al., 2012; Sommers, et al., 2015). The present study was conducted with German stimuli and listeners. Overall, the results between the two languages are similar, which aligns with our expectations. Similarly, the differences between vowels and plosives in terms of their length and intensity of the signal, do not differ between languages either, so we would expect similar results in other languages that are tested.

The present study compared white noise and babble noise. The babble noise used came from a recording of café noise with multiple speakers and some level of environmental noise. As there were multiple speakers, there was a larger difference to the single-speaker target, due to the signal being less similar to a single speaker. We expect that if the experiment would be repeated with babble noise containing a low number of speakers (e.g., one or two), the difference between white noise and babble noise would be larger. Single-speaker competing speech, particularly in the same language and with a speaker of the same gender, would lead to more masking, both energetic and informational compared to the babble used in the current study.

Even though we already find high rates of mishearing in our study, it is likely that this underestimates the amount of mishearing that would occur for these materials in a more naturalistic setting. As shown by the learning effects, participants were aware of the possible semantic mismatches in the presented audio and sentence context. Therefore, they might have paid extra careful attention to the speech, more than they would have done in more natural circumstances. Analysis of the wrong responses showed that participants in fact paid a lot of attention to the acoustic signal rather than the sentence context.

Acknowledgments

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102.

References

Alm, M., Behne, D. M., Wang, Y., & Eg, R. (2009). Audio-visual identification of place of articulation and voicing in white and babble noise. *The Journal of the Acoustical Society of America, 126*(1), 377-387.

Alwan, A., Jiang, J., & Chen, W. (2011). Perception of place of articulation for plosives and fricatives in noise. *Speech Communication, 53*(2), 195-209.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Boersma, P. & Weenink, D. (2021): Praat: doing phonetics by computer [Computer program]. Version 6.1.05, retrieved from http://www.praat.org/

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America, 84*(1), 101-114.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101-1109.

Cooke, M. (2009). Discovering consistent word confusions in noise. In *Tenth Annual Conference of the International Speech Communication Association*.

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America, 107*(1), 538-546.

Edwards, T. J. (1981). Multiple features analysis of intervocalic English plosives. *The Journal of the Acoustical Society of America, 69*(2), 535-547.

Failes, E., Sommers, M. S., & Jacoby, L. L. (2020). Blurring past and present: Using false memory to better understand false hearing in young and older adults. *Memory & Cognition, 48*(8), 1403-1416.

Flanagan, J. L. (1955). A Difference Limen for Vowel Formant Frequency. *The Journal of the Acoustical Society of America, 27*(3), 613–617.

Gordon-Salant, S. (1985). Some perceptual properties of consonants in multitalker babble. *Perception & psychophysics, 38*(1), 81-90.

Hutchinson, K. M. (1989). Influence of sentence context on speech perception in young and older adults. *Journal of Gerontology, 44*(2), P36-P44.

Krech, E., Stock, E., Hirschfeld, U., & Anders, L. (2009). *Deutsches Ausspracheworterbuch*. Berlin, Boston: De Gruyter Mouton.

Lecumberri, M. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119*(4), 2445-2454.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied, 68*(8), 1-13.

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. (2016). Panphon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers,* 3475-3484.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America, 87*(6), 2705-2715.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97*(1), 593-608.

R Development Core Team. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://wwwR-project.org/

Rogers, C. S. (2017). Semantic priming, not repetition priming, is to blame for false hearing. *Psychonomic Bulletin & Review, 24*(4), 1194-1204.

Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: the role of age differences in metacognition. P*sychology and Aging, 27*(1), 33-45.

Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing?. *Psychology and Aging, 20*(2), 261-271.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America, 118*(5), 2775-2778.

Soli, S. D., & Arabie, P. (1979). Auditory versus phonetic accounts of observed confusions between consonant phonemes. *The Journal of the Acoustical Society of America, 66*(1), 46-59.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychology and Aging, 14*(3), 458-472.

Sommers, M. S., Morton, J., & Rogers, C. (2015). You are not listening to what I said: False hearing in young and older adults. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger III (Eds.), *Remembering: Attributions, processes, and control in human memory (essays in Honor of Larry Jacoby)* (pp. 269–284). New York, NY: Psychology Press.

Taitelbaum-Swead, R., & Fostick, L. (2016). The effect of age and type of noise on speech perception under conditions of changing context and noise levels. *Folia Phoniatrica et Logopaedica, 68*(1), 16-21.

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science, 14*(3), 144-148.

Wingfield, A., Tun, P. A., & Rosen, M. J. (1995). Age differences in veridical and reconstructive recall of syntactically and randomly segmented speech. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 50B*(5), P257-P266.