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

Marjolein van Os, Jutta Kray, Vera Demberg (Saarland University)

Language understanding is facilitated by highly predictive contexts even in noisy conditions (Dubno, Horwitz, & Ahlstrom, 2000; Sommers & Danielson, 1999). Here we investigated whether the type of noise influences speech comprehension, that is, the recognition of minimal pairs, differently. While multi-speaker babble noise approximates the average long-term spectrum of the speech of an adult male speaker, white noise has a flat spectral density with the same amplitude throughout the audible frequency range. Both types of noise lead to energetic masking of the target speech, as both the speech signal and the noise have energy in the same spectral frequency bands (Brungart, 2001). However, as babble noise shows more overlap with the spectral information of a single speaker, it may lead to greater energetic masking than the more spread out energy of white noise. While previous studies have compared babble and white noise (Lecumberri & Cooke, 2006; Taitelbaum-Swead & Fostick, 2016), no studies have so far directly compared the effect of noise type on mishearing in different noise contexts. We expected that babble noise reduces recognition performance more than white noise and that this effect is more pronounced when sentence endings are unpredictable target words.

To examine this, participants listened to recordings of sentences embedded in babble and white noise at -5 dB SNR and in quiet. They typed in the last word of the sentence they had heard. Each sentence was presented visually up to the target word to provide a predictive context that was understandable regardless of background noise. The final target word either fit the sentence semantically (mean cloze 0.72, high predictability condition, HP), or was unpredictable based on the preceding context (mean cloze 0, low predictability condition, LP). Example stimuli can be found in Table 1. The target words formed minimal pairs differing in one phonetic feature in medial position, and were swapped with the respective sentence frames to create LP items. This allowed us to investigate whether listeners can rely on small acoustic cues for word recognition, even in noise, while keeping sentence contexts equal across conditions.

Responses from 48 participants (31 males, mean age = 24 years) 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*). Using a General Linear Mixed Model with fixed effects of Noise and Predictability as well as the interaction, and Trial No, and random intercepts for Subject and Item, with random intercepts for Noise and Predictability for both, we find that both noise conditions lead to fewer correctly identified target responses than quiet ($\beta = -5.30$, $SE = 0.85$, $z = -6.21$, $p < .001$ for babble and $\beta = -4.50$, $SE = 0.82$, $z = -5.51$, $p < .001$ for white noise). The rate of correctly identified targets does not differ significantly between the two noise conditions ($p = .09$). Regarding the beneficial effect of predictability, we find that participants correctly identify the target more often for HP compared to LP ($\beta = -6.01$, $SE = 0.91$, $z = -6.58$, $p < .001$; see Figure 1). On the subset of unpredictable items, we next tested whether the types of errors (wrong vs. distractor) differ between the noise conditions, see Figure 2. Here the distractor fit the context and most of the acoustic signal. The wrong response did not fit both. We ran the model with fixed effects of Noise and Trial No, and random intercepts for Subject and Item. We find more wrong responses in babble noise compared to quiet ($\beta = -1.23$, $SE = 0.34$, $z = –3.61$, $p < .001$), as well as a to white noise ($\beta = 0.73$, $SE = 0.24$, $z = 3.01$, $p < .01$). 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.

The results suggest that noise hamper speech comprehension irrespective of sentence predictability. The type of noise induced different errors indicating that white noise is indeed an easier condition than babble. Analyses of semantic fit and phonetic distance in the wrong responses will shed more light on this.

Table 1. Example Stimuli

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

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 have been given in *italics.*

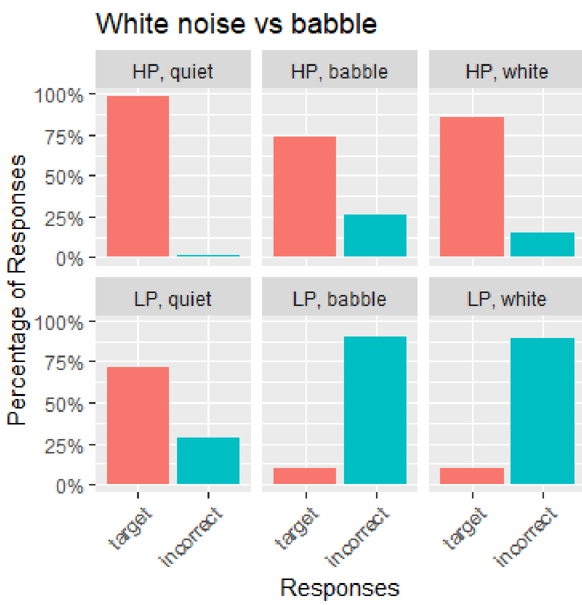

Figure 1: % of target and incorrect responses for high (HP) and low predictability condition (LP) in quiet, babble noise, and white noise.

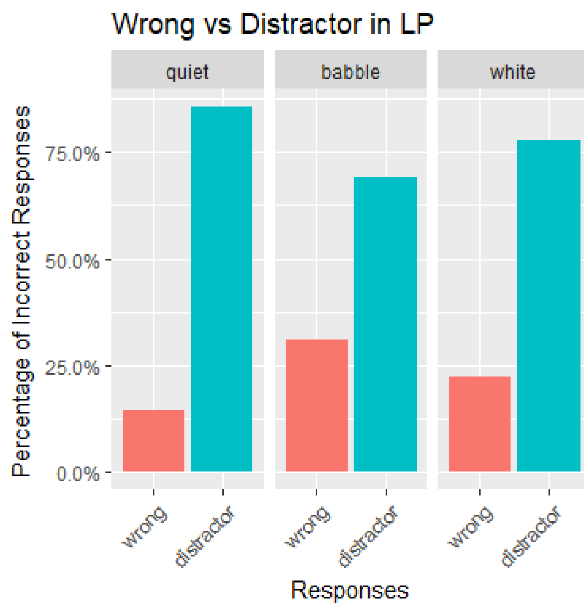

Figure 2: % of wrong and distractor responses for the low predictability condition (LP) in quiet, babble noise, and white noise.

References

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.

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.

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.

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.

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.