Interspeech 2024
1-5 September 2024, Kos, Greece

Towards a better understanding of receptive multilingualism: listening conditions and priming effects

Wei Xue[1], *Ivan Yuen*[1], *Bernd Möbius*[1]

[1]Language Science and Techonology, Saarland University, Germany

weixue@lst.uni-saarland.de, ivyuen@coli.uni-saarland.de, moebius@lst.uni-saarland.de

Abstract

Receptive multilingualism is a form of communication where speakers can comprehend an utterance of a foreign language (Lx) using their native language (L1) when L1 and Lx share similarities in, e.g., vocabulary and pronunciation. The success of receptive multilingualism can be tested by examining accuracy and reaction time of auditory word recognition (AWR) of target words in lexical decision tasks. AWR in such tasks can be affected by adverse listening conditions due to environmental noises and by the presence of a preceding prime word. This study explores whether AWR of L1 in Lx-L1 pairs (Lx = Dutch; L1 = German or English) will be affected by different degrees of similarities in their phonology and semantics and whether such an influence will differ as a function of listening condition. We observed less accurate and slower responses without semantic similarity but a null effect on accuracy without phonological overlap. The interaction with listening conditions is language-dependent.

Index Terms: receptive multilingualism, intelligibility, auditory word recognition, lexical decision task, cognate

1. Introduction

Receptive multilingualism is a form of communication where speakers can comprehend an utterance of a foreign language (Lx) using their native language (L1) when L1 and Lx share similarities in, e.g., vocabulary or pronunciation or both [1, 2]. In this case, the two languages are considered to be mutually intelligible [3]. Studying receptive multilingualism can help understand communication barriers and linguistic diversity in general.

Receptive multilingualism can be studied through various tasks such as opinion testing and lexical decision tasks (see [4] for an overview). In a lexical decision task, researchers examined the comprehension performance in terms of auditory word recognition (AWR) as indexed by reaction time (RT) [5] and accuracy [6]. AWR requires listeners to decode and map acoustic-phonetic patterns or representations to some semantic units [7].

However, previous studies are limited. They rarely examined receptive multilingualism in a realistic listening situation with environmental noise. These issues deserve attention, given the recent findings that noisy environments can affect people's speech production [8] and perception [9, 10], including AWR.

The recognition of auditory words can also be influenced by an acoustic prime and the relation between prime-target pairs, i.e., whether they share similarities in their phonology and semantics. Regarding prime-target pairs with overlap in phonological form, previous studies have reported facilitatory effects on the speed of recognition in native languages [11] or in the context of bilingualism [6, 12]. Such facilitatory effects are known as phonological priming effects. Regarding prime-target pairs with overlaps in semantics, previous research has reported shorter RT when target words were preceded by semantically related word primes than by unrelated word primes in cross-language context [6, 12, 13, 14]. This phenomenon is known as semantic priming [13].

Taking together the research on speech production and perception in adverse listening conditions and the research on AWR with priming effects, the following two research questions remain unanswered:

- RQ1: Will AWR of target words in L1 be primed by degrees of phonological or semantic similarity in Lx-L1 prime-target pairs?
- RQ2: Will such an influence differ as a function of listening condition?

To address RQ1, we introduced cognates (CG), which overlap in orthographic-phonological form across two languages and share meaning, as the baseline[1]. We compared CG with two other types: phonological false friends (FF), which overlap in orthographic-phonological form across two languages but have different meanings, and translation equivalents (TE), which have the same meaning but deviate in orthographic-phonological form across two languages[2]. The aim of comparing CG vs. FF and CG vs. TE was to explore whether the lack of semantic similarity and phonological overlap, respectively, would negatively affect AWR.

To address RQ2, we studied prime-target pairs in five listening conditions: (1) quiet, (2) white noise (WN) with Signal-to-Noise Ratio (SNR) being 0 dB (WN0), (3) white noise with SNR = -6 dB (WN-6), (4) babble noise (BN) with SNR = 0 dB (BN0), and (5) babble noise with SNR = -6 dB (BN-6).

In the current study, Dutch served as the Lx for the prime words, and German and English as the target L1 languages. These three languages are from the same, Germanic, language family and posited to be mutually intelligible to some extent. Therefore, the two research questions were tested in Dutch-German and Dutch-English experiments separately.

2. Method

2.1. Prime-target pairs

The above-mentioned three word-pair types were used as stimuli: CG, FF, and TE. We also included non-words as fillers. There were 36 CG, 27 FF, 33 TE, and 30 fillers in the Dutch-German experiment. The Dutch-English experiment contained 36 CG, 38 FF, 48 TE, and 30 fillers. All the Dutch, German, and

[1]An Example of CG for Dutch-English is *contact-contact*.

[2]Examples of FF and TE for Dutch-English are *wet-wet*, where *wet* in Dutch means 'law' in English, and *huur-rent*, respectively

 10.21437/Interspeech.2024-418

English words of the CG type were selected from the datasets of cognates in [15] and in [16]. These selected words have word lemma frequencies larger than 20 (per million) in CELEX [17] and 10 (per million) in SUBTLEX [18, 19, 20]. Since it was difficult to find enough proper FF and TE pairs, their word frequencies were not controlled. It is also worth noting that the number of test pairs varied in the Dutch-German and the Dutch-English experiments because German and English differ in their similarity to Dutch. Therefore, it is difficult to have as many Dutch-German pairs as Dutch-English pairs. This also explains why we employed different numbers of pairs across four word-pair types.

2.2. Recordings of prime-target pairs

We recruited three female native speakers (mean age=27.3, SD=4.0), one for each language. They were instructed to read aloud the words in a sound-attenuated booth. To elicit the Lombard effect, the speakers were exposed to noise through headphones. The noise level was presented at a sound pressure level of 70 dB. Depending on the listening condition, either white noise or café noise from the BBC Sound Effects Library[3] were used. The speakers were not informed about what type of words they were reading, and the order of the prime-target pairs was randomized. They read each word three times to allow the authors to choose a proper candidate. During the recording, they sat still on a chair, and the words were shown five at a time on a screen in front of them. A standing microphone was placed at a distance of 1 meter from the speakers to prevent headphone noise from being recorded. Each word was extracted from the recordings with intensity rescaled to the same level at 70 dB. These rescaled words were merged with the corresponding noise types (white or babble noise) at the required levels (i.e., SNR=0 dB and SNR=-6 dB)[4] using a custom Praat script [21]. We added 500 ms of noise before the onset and after the offset of each word to familiarize participants with the noise. We also added 500 ms of silence for speech in the quiet condition to minimize the difference in the settings of listening conditions other than noise types and noise levels. All speakers gave their consent verbally before the recording procedure.

2.3. Experiment implementation

Both the Dutch-German and Dutch-English experiments were set up using Labvanced [22]. In each experiment, there were five separate sessions for the five listening conditions in the order of (1) WN0, (2) WN-6, (3) quiet, (4) BN0, and (5) BN-6. The blocked design was used to reduce participants' cognitive load arising from switching between noise and word type conditions. The prime-target pairs were randomized across five sessions, such that every pair appeared only once throughout the experiment. This design was aimed to minimize repetition effects [6]. This implementation also means that each session had a subset of the prime-target pairs, namely a subset of 7 CG, 5 FF, 6 TE, and 6 fillers for the Dutch-German experiment and 7 CG, 7 FF, 9 TE, and 6 fillers for the Dutch-English experiment. In each session, participants first went through four practice trials and then proceeded to the test trials. In each trial, the recordings of a prime-target pair were displayed automatically. Participants were instructed to decide if the second, target word is a real, existing word in German/English and react as accurately and quickly as possible by pressing key 'D' for a real word and key 'K' for a nonce word. We collected the key-press responses and the RT from the prime word onset to the key-pressing through Labvanced.

[3] Crowds: Interior, Dinner-Dance, https://bbcsfx.acropolis.org.uk/

[4] Merging was accomplished by preserving speech and modifying noise

2.4. Participants

We recruited 84 participants via Prolific[5], 42 (mean age=29.5, SD=5.0) for the German and 42 (mean age=32.2, SD=5.1) for the English experiment. The participants were gender-balanced. Most (86%) of the German participants were raised monolingually in German. 76% spoke at least one other language such as English at a self-reported advanced level, and only two of them spoke limited Dutch. Some of them spoke another language other than English or Dutch (e.g., French, Russian, and Spanish) at a beginner or intermediate level, but none of these languages were Germanic. All English participants were located in the United States and 93% were monolingual English speakers. Two of them spoke another language but did not report which language. None of the participants reported any hearing loss or attention-deficit/hyperactivity disorder. All participants gave informed consent before they started the experiment. The studies were approved by the ethics committee of Saarland University.

2.5. Statistical analyses

To evaluate the performance in AWR, we considered two metrics: (1) accuracy in terms of the responses being correct or incorrect and (2) absolute RT (abRT), calculated as the temporal difference from the onset of the target word to the key-press response, after subtracting the duration of the auditory prime and the 500 ms of silence or noise before the target word. We only considered abRT for the items that were correctly responded to. We further excluded responses to filler words and those outside of the abRT range (150 ms $<$ abRT $<$ 3000 ms). The floor threshold of 150 ms was taken to exclude accidental responses [5], whereas the ceiling threshold of 3000 ms was defined to exclude responses to which participants did not pay attention.

To address the two research questions, we used Treatment contrast coding[6] and specifically analyzed the difference between the following contrasts. For RQ1 regarding the effect of word type, we considered CG as the baseline and examined two contrasts. The first contrast (i.e., CG vs. FF) showed the difference between CG and FF (with weights of 0, -1, 0, presenting in the order of CG, FF, and TE). The second contrast (i.e., CG vs. TE) showed the difference between CG and TE (weights: 0, 0, -1). The contrasts tested the null hypotheses of $mean_{CG} - mean_{FF} = 0$ and $mean_{CG} - mean_{TE} = 0$, respectively. For RQ2 regarding listening conditions, we considered the quiet condition as the baseline and examined four contrasts. The first contrast (i.e., Quiet vs. Noise) showed the difference between the quiet condition and the mean of all four listening conditions with noise (weights: -1, -1, 0, -1, -1, presenting in the order of the five sessions). This contrast tested the null hypothesis of $mean_{Quiet} - mean_{Noise} = 0$. The second contrast (i.e., WN vs. BN) showed the difference between white noise and babble noise (weights: 0.5, 0.5, 0, -0.5, -0.5,) and tested the null hypothesis of $mean_{WN} - mean_{BN} = 0$. The third and fourth contrasts (i.e., WN-6 vs. WN0 and BN-6 vs. BN0) showed the difference between SNR=-6 dB and SNR=0 dB nested within white noise (weights: -0.5, 0.5, 0, 0, 0) and

[5] https://www.prolific.com

[6] We followed the instruction in [23].

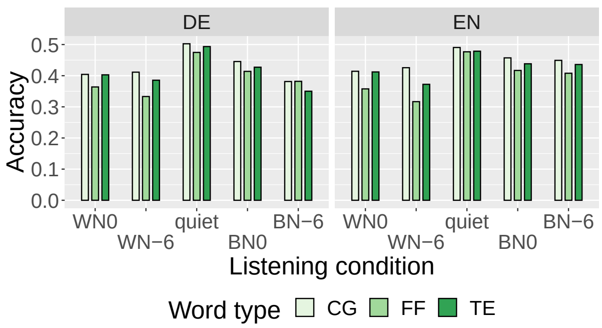

Figure 1: *Accuracy (in proportion) for German (DE) and English (EN) experiments.*

babble noise (weights: 0, 0, 0, -0.5, 0.5). These contrasts tested the null hypotheses of $mean_{WN-6} - mean_{WN0} = 0$ and $mean_{BN-6} - mean_{BN0} = 0$, respectively.

We applied generalized linear mixed models (GLMMs) to predict accuracy and linear mixed models (LMMs) to predict abRT by the above-mentioned contrasts and with[7] or without interactions[8] between them. We modelled participants and prime-target pair items as random effects for intercept. All GLMMs were run with a logit link and were controlled with the bobyqa optimizer and a maximum number of iterations of $2*10^5$ to improve model convergence. We applied the Akaike Information Criterion (AIC) to select a final model from those in which there were with or without interactions as mentioned above. We only report the results of the final models that had lower AIC values in this paper. All statistical analyses were conducted in R [24] by using the lme4 package [25] for the GLMMs and lmerTest [26] for the LMMs, as well as ggplot2 [27] for visualization.

3. Results

3.1. Descriptive statistics

A visualization of accuracy and abRT is shown in Figures 1 and 2, respectively. As shown in Figure 1, the accuracy (in proportion) of CG is higher than those of FF and TE, irrespective of the L1 language. The accuracy values are generally lower in conditions with noise and also generally lower in WN compared to BN conditions. The distributions of abRT for German and English are similar, as shown in Figure 2. Furthermore, abRT values are lower for CG than for FF and TE in the quiet condition and higher in conditions with noise, as shown in Figure 3.

3.2. Predicting accuracy with word types and listening conditions

We applied GLMMs to predict accuracy (i.e., a binomial variable of 'correct' and 'incorrect') with the seven contrasts explained in section 2.5. The estimates of intercept are the mean values of CG in the quiet condition, i.e., the baseline.

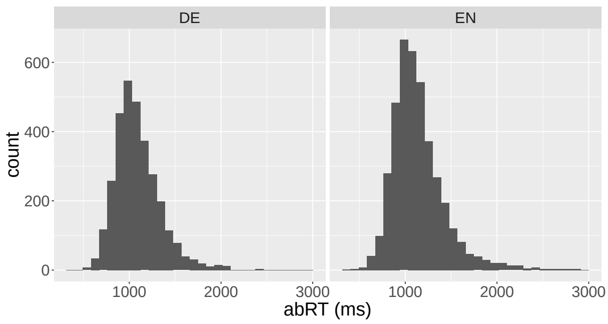

Figure 2: *Histogram of abRT (ms) for German (DE) and English (EN) experiments.*

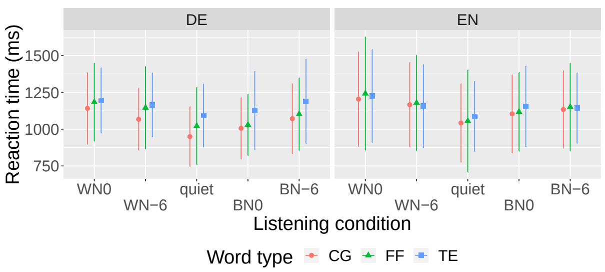

Figure 3: *Mean and error bars of abRT (ms) in German (DE) and English (EN) experiments with the five listening conditions.*

As the Quiet vs. Noise contrast tested the null hypothesis of $mean_{Quiet} - mean_{Noise} = 0$, positive estimates ($\beta$) indicate that the Quiet condition has higher values (i.e., $mean_{Quiet} - mean_{Noise} > 0$) and vice versa. Likewise, the contrast of WN vs. BN tested the null hypothesis of $mean_{WN} - mean_{BN} = 0$, and thus positive estimates indicate that white noise conditions have higher values (i.e., $mean_{WN} - mean_{BN} > 0$) and vice versa. Similarly, the CG vs. FF tested the null hypothesis of $mean_{CG} - mean_{FF} = 0$, and positive estimates indicate that the CG has higher values compared to FF (i.e., $mean_{CG} - mean_{FF} > 0$). This also applies to interpreting the contrasts between noise levels and for CG vs. TE.

3.2.1. Dutch-German experiment

The best model fit identified the GLMM with interactions. The results revealed a significant effect (i.e., $p < 0.05$) of CG vs. FF (β = 2.375, *SE* = 0.797, z = 2.979, p <0.01) but no significant effect of CG vs. TE (β = 1.398, *SE* = 0.845, z = 1.655, p=0.098). The Quiet vs. Noise contrast comparison also reached statistical significance (β = 3.68578, *SE* = 0.706, z = 5.223, p<0.001) as well as BN-6 vs. BN0 (β = -1.01453, *SE* = 0.240, z = -4.234, p<0.001). We also found a significant interaction between CG vs. FF and Quiet vs. Noise (β = 1.80511, *SE* = 0.769, z = 2.347, p<0.05), indicating that the difference between CG and FF is larger in Quiet compared to in Noise.

3.2.2. Dutch-English experiment

The best model fit identified the GLMM without interactions, indicating that the two types of contrasts contribute individually. The results revealed a significant effect of CG vs. FF (β = 1.200, *SE* = 0.323, z = 3.717, p<0.001) but no significant effect of CG vs. TE (β = 0.5664, *SE* = 0.310, z = 1.826, p = 0.068).

[7] As the dependent variables (DV) for GLMMs and LMMs were accuracy and abRT, the model with interaction is DV ∼ CG vs. FF * Quiet vs. Noise + CG vs. FF * WN vs. BN + CG vs. FF * BN-6 vs. BN0 + CG vs. FF * WN-6 vs. WN0 + CG vs. TE * Quiet vs. Noise + CG vs. TE * WN vs. BN + CG vs. TE * BN-6 vs. BN0 + CG vs. FF * WN-6 vs. WN0 + (1|item) + (1|participant).

[8] The model without interaction is DV ∼ CG vs. FF + CG vs. TE + Quiet vs. Noise + WN vs. BN + BN-6 vs. BN0 + WN-6 vs. WN0 + (1|item) + (1|participant).

The Quiet vs. Noise contrast comparison also reached statistical significance (β = 2.3395, *SE* = 0.223, *z* = 10.513, $p<0.001$). These patterns are similar to those of the Dutch-German experiment. Unlike the results for L1 German speakers, we found significant effects of WN vs. BN (β = -0.9733, *SE* = 0.102, *z* = -9.530, $p<0.001$) as well as WN-6 vs. WN0 (β =-0.3505, *SE* = 0.128, *z* = -2.749, $p<0.01$).

3.3. Predicting abRT with word types and listening conditions

We applied LMMs for predicting abRT (i.e., a continuous variable) with the seven contrasts explained in section 2.5. Similar to the explanation above, positive estimates indicate higher values for Quiet than for Noise in the contrast of Quiet vs. Noise and higher values for WN than for BN in the contrast of WN vs. BN. This also applies for interpreting the contrasts between noise levels. Positive estimates for CG vs. FF and CG vs. TE indicate that CG has higher values compared to FF and TE.

3.3.1. Dutch-German experiment

The best model fit identified the LMM with interactions. The results revealed significant effects of CG vs. FF (β = -67.651, *SE* = 22.485, *t* = -3.009, $p<0.01$) and CG vs. TE (β = -141.401, *SE* = 21.222, *t* = -6.663, $p<0.001$). We found significant effects for all four contrasts regarding listening conditions: Quiet vs. Noise (β = -117.224, *SE* = 11.618, *t* =-10.090, p<0.001), WN vs. BN (β = 69.718, *SE* = 11.296, *t* = 6.172, $p<0.001$), WN-6 vs. WN0 (β = -75.955, *SE* = 15.945, *t* = -4.764, $p<0.001$), and BN-6 vs. BN0 (β = 69.997, *SE* = 15.875, *t* = 4.409, $p<0.001$). The interaction between CG vs. TE and Quiet vs. Noise reached statistical significance (β =-36.897, *SE* = 17.248, *t* = -2.139, p=0.033), indicating that the difference between CG and TE is larger in Quiet compared to in Noise. We also observed a significant interaction between CG vs. TE and WN vs. BN (β = 43.249, *SE* = 16.894, *t* = 2.560, p=0.011) as well as between CG vs. TE and WN-6 vs. WN0 (β = -46.745, *SE* = 23.646, *t* = -1.977, p=0.048).

3.3.2. Dutch-English experiment

The best model fit identified the LMM with interactions. The results revealed a significant effect of CG vs. TE (β = -50.908, *SE* = 22.853, *t* = -2.228, p=0.027) but no significant effect of CG vs. FF (β = -27.254, *SE* = 24.191, *t* = -1.127, p=0.2608). We found significant effects of three contrasts regarding listening conditions: Quiet vs. Noise (β = -120.442, *SE* = 13.709, *t* = -8.786, $p<0.001$), WN vs. BN (β = 66.447, *SE* = 12.927, *t* = 5.140, $p<0.001$), and WN-6 vs. WN0 (β = -38.930, *SE* = 18.441, *t* = -2.111, p=0.035). No significant results were found for the interactions indicating that the effects of CG vs. FF and CG vs. TE are similar in Quiet vs. Noise, WN vs. BN, and for the contrasts between SNR=0 dB and SNR=-6 dB.

4. Discussion and conclusion

In this study, we explored the effects of word type and listening condition on receptive multilingualism. For the effect of word type, we examined whether the accuracy and reaction time (RT) of target L1 words in Lx prime–L1 target pairs would be influenced by different degrees of semantic and/or phonological similarity, by comparing cognates (CG) with phonological false friends (FF) and translation equivalents (TE). We further examined whether such an influence differs as a function of the listening condition.

Our analyses revealed that without semantic similarity (CG vs. FF), responses to the target words are significantly less accurate for both L1 German and English speakers, but only slower in RT for the L1 German speakers. This result seems to partially align with previous research where responses to semantically related words were found to be faster than those to unrelated word primes [6, 12, 13, 14, 5]. The reason for the slow response from L1 German speakers could be that more than half of the German speakers also spoke advanced English. Perhaps this increases the size of potential FF, which may lead to phonological ambiguity and thus slower responses. As such, there is a limitation of the present study due to the different linguistic backgrounds of the L1 German and English participants. It is worth analyzing further how mastering another related language (e.g., English in this study) affects participants' responses to L1 in Lx-L1 priming.

In contrast, the lack of phonological similarity (CG vs. TE) does not significantly affect accuracy, although it does affect RT. The null effect on accuracy is in line with the findings in [6] that phonological cognates and non-cognates resulted in comparable accuracy in noise when they were preceded by a semantically related prime. The observation of slow reaction time for TE in the current experiment might be due to the absence of any phonological 'anchor'. Translation equivalents between Lx and L1 are based on word meaning. Without any phonological similarity in form, it is difficult to make an 'educated' guess, because the auditory input from the prime in Lx might activate a larger set of possible word options in L1 and therefore lead the listeners away from the L1 target.

Our analysis across listening conditions revealed that the presence of noise tends to significantly reduce the accuracy and slow down the responses for both L1 German and English speakers, which is in line with the effect of noise in speech perception for both native and non-native studies [9, 10]. White noise tends to reduce accuracy only for the L1 English speakers and slow down the RT for both L1 German and English speakers. Noise level under white noise significantly decreased accuracy for the L1 English speakers but slightly sped up RT for both L1 German and English speakers. However, only L1 German speakers decreased their response accuracy and slowed down RT under babble noise. In addition, the significant effect of word type differs as a function of the listening condition only for the L1 German speakers, implying a language-dependent interaction. Therefore, a limitation of this study is that we did not control for word frequency and other linguistic characteristics of FF and TE. Future research should further analyze and interpret the interaction in detail, as well as the different results between languages, by taking into account linguistic factors such as phonetic distance and word adaptation surprisal [5] between prime-target pairs. Also, it is interesting to explore whether the results would remain the same if we have recordings of male native speakers from the three languages.

In conclusion, false friends, reflecting a lack of semantic similarity, seem to have an attenuating effect on accuracy and RT in receptive multilingualism, underscoring the role of phonological ambiguity and potential misperception. The presence of noise affects the response in auditory word recognition but its interaction with word type seems to be language-dependent. Our study sheds light on a better understanding of receptive multilingualism.

5. Acknowledgements

This research is supported by a Rubicon grant from the Dutch Research Council (NWO) to the first author (019.223SG.004) and a grant by German Research Foundation (DFG, Project ID 232722074-SFB1102) to the second and third authors.

6. References

[1] R. Beerkens, *Receptive multilingualism as a language mode in the Dutch-German border area*. Münster: Waxmann, 2010.

[2] J. Kudera, *Slavic receptive multilingualism: intercomprehension of speech [Doctoral dissertation]*. Saarländische Universitäts- und Landesbibliothek, 2022.

[3] C. Gooskens, V. J. Van Heuven, J. Golubović, A. Schüppert, F. Swarte, and S. Voigt, "Mutual intelligibility between closely related languages in europe," *International Journal of Multilingualism*, vol. 15, no. 2, pp. 169–193, 2018.

[4] C. Gooskens, *Experimental methods for measuring intelligibility of closely related language varieties*. Oxford University Press, New York, 2013, pp. 195–213.

[5] J. Kudera, P. Georgis, B. Möbius, T. Avgustinova, and D. Klakow, "Phonetic distance and surprisal in multilingual priming: Evidence from slavic," in *Proc. INTERSPEECH 2021 – 22rd Annual Conference of the International Speech Communication Association*, 2021, pp. 3944–3948.

[6] S. Guediche, M. Baart, and A. G. Samuel, "Semantic priming effects can be modulated by crosslinguistic interactions during second-language auditory word recognition," *Bilingualism: Language and Cognition*, p. 1–11, 2020.

[7] A. Weber and O. Scharenborg, "Models of spoken-word recognition," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 3, no. 3, pp. 387—-401, 2012.

[8] O. Ibrahim, I. Yuen, M. van Os, B. Andreeva, and B. Möbius, "The combined effects of contextual predictability and noise on the acoustic realisation of german syllables," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 911–920, 2022.

[9] O. Scharenborg and M. van Os, "Why listening in background noise is harder in a non-native language than in a native language: A review," *Speech Communication*, vol. 108, no. 2, pp. 53–64, 2019.

[10] S. Varley, *Bilingual speech perception in noise: effects of target predictability, noise type, and language background [Doctoral dissertation]*. University of Pittsburgh, 2020.

[11] J. G. van Hell and T. Dijkstra, "Foreign language knowledge can influence native language performance in exclusively native contexts," *Psychonomic Bulletin Review*, vol. 9, no. 4, pp. 780—-789, 2002.

[12] I. G. Temnikova and O. V. Nagel, "Effects of cognate and relatedness status on word recognition in russian-english bilinguals of upper-intermediate and advanced levels," *PProcedia-Social and Behavioral Sciences*, vol. 200, pp. 381–386, 2015.

[13] K. Rataj, P. Kakuba, P. Mandera, and W. J. B. van Heuven, "Establishing semantic relatedness through ratings, reaction times, and semantic vectors: A database in polish," *PLoS ONE*, vol. 18, no. 4, p. e0284801, 2023.

[14] M. Perea, J. A. Dunabeitia, and M. Carreiras, "Masked associative/semantic priming effects across languages with highly proficient bilinguals," *Journal of Memory and Language*, vol. 58, no. 4, pp. 916–930, 2008.

[15] J. Schepens, T. Dijkstra, F. Grootjen, and W. J. B. Van Heuven, "Cross-language distributions of high frequency and phonetically similar cognates," *PloS one*, vol. 8, no. 5, 2013.

[16] E. D. Poort and J. M. Rodd, "A database of dutch-english cognates, interlingual homographs and translation equivalents," *Journal of Cognition*, vol. 2, no. 1, 2019.

[17] R. H. Baayen, R. Piepenbrock, and L. Gulikers, "Celex2 ldc96l14 [web download]," 1995.

[18] E. Keuleers, M. Brysbaert, and B. New, "Subtlex-nl: a new measure for dutch word frequency based on film subtitles," *Behavior Research Methods*, no. 3, 2010.

[19] M. Brysbaert, M. Buchmeier, M. Conrad, A. M. Jacobs, J. Bölte, and A. Böhl, "The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in german," *Experimental Psychology*, no. 5, 2011.

[20] D. A. Balota, M. J. Yap, M. J. Cortese, K. A. Hutchison, B. Kessler, B. Loftis, J. H. Neely, D. L. Nelson, G. B. Simpson, and R. Treiman, "The english lexicon project," *Behavior Research Methods*, vol. 39, 2007.

[21] M. Winn, "Mix_speech_and_noise.txt." [Online]. Available: https://github.com/ListenLab/Praat/blob/master/Mix_speech_and_noise.txt

[22] H. Finger, C. Goeke, D. Diekamp, K. Standvoß, and P. König, "Labvanced: a unified javascript framework for online studies," 2017.

[23] D. J. Schad, S. Vasishth, S. Hohenstein, and R. Kliegl, "How to capitalize on a priori contrasts in linear (mixed) models: A tutorial," *Journal of memory and language*, vol. 110, 2020.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023. [Online]. Available: https://www.R-project.org/

[25] D. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[26] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[27] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. Available: https://ggplot2.tidyverse.org