# Masker Language and Acoustic Confusability: Effects on Letter Sequence Recognition

*Jessica L. L. Chin, Laurence Bruggeman, Mark Antoniou*

The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University

jessica.chin/l.bruggeman/m.antoniou@westernsydney.edu.au

## Abstract

Speech-in-speech recognition is harder when the target and masker language are the same, and when cognitive load increases, but it is unclear whether the listener's language experience and the masker's similarity to the target also have an effect. Here, English monolinguals and Arabic–English bilinguals recalled English letter sequences with low or high acoustic confusability and embedded in masker speech. Results showed that a masker in the language of the target speech and high confusability are most detrimental to speech recognition. The Swedish masker's similarity to English did not affect recognition, nor did the bilinguals' familiarity to the Arabic masker.

**Index Terms**: speech-in-speech recognition, linguistic release from masking, linguistic similarity hypothesis, cognitive load

## 1. Introduction

Listeners can understand speech even in suboptimal conditions, such as noisy environments. They attend to a talker's speech (the target) whilst inhibiting distracting speech or other sounds (the masker) in a phenomenon commonly known as the "cocktail party effect" [1], [2]. If the masker becomes too distracting, recognising the target speech becomes hard or even impossible. We distinguish two main types of masking: energetic and informational. During energetic masking, the spectral qualities of the masker signal inhibit the intelligibility of the target speech. Informational masking refers to the additional aspects of a signal (outside of energetic masking) that can hinder target recognition, such as linguistic features in masker speech [1]. For instance, when listening to an English target, an English masker might make recognition more difficult than a Dutch masker due to the maskers' linguistic differences.

Some common findings arise in speech-in-speech recognition research. For instance, the more typologically similar a masker is to the target, the more difficult speech recognition becomes. This has been observed in multiple studies where the target and masker are the same language, e.g., English-in-English [3]–[6]. However, results have been varied for conditions where the target and masker languages differ, but are typologically similar. While maskers with the same rhythmic structure as the target (e.g., stress-, syllable-, or mora-timed) have been shown to be more detrimental to speech recognition than rhythmically dissimilar maskers [7], a recent study showed no significant differences in speech recognition accuracy between rhythmically similar or dissimilar maskers [8]. Indeed, rhythmic structure is only one element in which languages may share similarities, and similarities in other typological features would be worth exploring.

When accounting for the listener's language experience and their familiarity to a masker in speech-in-speech recognition, results have also been varied. Some studies have found that nonnative (L2) listeners are poorer at recognising speech within speech than native (L1) listeners [9], [10]. When the masker language is known to the L2 listeners but not the L1 listeners (e.g., English-in-Mandarin), L2 listeners perform worse in this condition [11]. This suggests that listeners encounter difficulties when they know the masker language, even when the target and masker language are mismatched. When comparing between more proficient listeners (i.e., late and early bilinguals), speech-in-speech recognition performance also varies. For instance, Spanish–English bilinguals who acquired English past age 14 found it harder to recognise English-in-English speech than English monolinguals and Spanish–English bilinguals who acquired English before age 6 [12]. Another study, however, showed that early Spanish–English bilinguals performed worse in English sentence recognition than monolinguals [13]. In contrast to the above findings, a study comparing English monolinguals and English–Greek bilinguals show comparable results between both groups for English targets in a Greek masker: a masker native to the bilinguals, yet foreign to the monolinguals [14]. In sum, a listener's language experience and their knowledge of a masker (that is not the same as the target language) can influence their speech-in-speech recognition performance, but there is no consensus for these patterns within the literature.

Speech-in-speech recognition typically requires more cognitive resources than listening to speech in quiet, as listening to a target while trying to ignore a masker requires attentional effort [15]. When considering the mechanisms involved in processing speech, the phonological loop allows for the short-term retention of verbal information, which also includes novel speech input [16]. There is also a link between acoustic representations of information and memory span, in which even visual representations of acoustically confusable letter sequences (e.g., BCD…) may lead to poorer retention scores than acoustically different (e.g., QYZ…) sequences [17]. Similar to the dividing of attention between multiple speech streams, the demand on working memory arising from implementing a secondary task can also make speech recognition more difficult [15]. In the current study, cognitive load, in the form of acoustical confusability of the target letter sequence, was manipulated to examine its effects on speech-in-speech recognition alongside masker language and the language experience of the listener.

The present study investigated the different linguistic and cognitive factors which influence speech-in-speech recognition. We aimed to determine how the phonetic similarity of the target and the masker languages affect speech-in-speech recognition. In addition, we were interested in whether the listener's knowledge of a masker language also impacts on

speech-in-speech recognition. Not much is known about how these factors interact with each other. Also unclear is whether these factors impact speech recognition to the same extent, or whether one factor perhaps has a greater impact than the other. The task we used was letter sequence recognition in masker speech. We also manipulated the cognitive load—by using sequences with either low or high acoustic confusability—to assess any further impacts on the listener's speech-in-speech recognition performance. This replicates a classic paradigm from Conrad and Hull [17], in which more acoustically confusable letter sequences were recalled with lower accuracy. These sequences, however, were presented visually. Our current study instead presents auditory letter sequences with varying levels of acoustic confusability, all in the presence of masker speech.

In sum, our research question is as follows: How is speech-in-speech recognition influenced by: 1) the phonetic similarity between the target and masker languages, 2) the listener's familiarity to the masker language, and 3) the cognitive load demands of the task?

To answer this question, we compared two groups of participants (Australian English monolinguals, Arabic–English bilinguals) on their performance recalling an auditory sequence of English letters. The sequences were presented within two-talker masker speech in one of four masker languages (Australian English, Arabic, Swedish, Spanish). Masker languages varied in their similarity to English (as measured by the size of their vowel inventory), as well as in their familiarity to the listener.

We predicted that the English masker would be most detrimental, as it is the same language as the target, and also the native language of all participants. We also predicted that the Arabic masker would pose more difficulty for the bilinguals than the monolinguals. If native knowledge of a masker is more detrimental to performance than phonetic similarity between the masker and target, we predict the following patterns: for the monolinguals, the English masker would be the most difficult condition, followed by Swedish, then Spanish or Arabic. For the bilinguals, English would also be the most difficult condition, followed by Arabic, then Swedish, then Spanish. Finally, we predicted that more acoustically confusable letter sequences would be more difficult to recall than acoustically different ones.

# 2. Method

## 2.1. Participants

Participants were 61 Australian English (AusE) monolinguals (43 females, 1 non-binary individual; $M_{Age}$ = 25.73, $SD$ = 7.95) and 20 Arabic–English bilinguals (17 females; $M_{Age}$ = 21.97, $SD$ = 7.18). Data from two bilingual participants were discarded for failure to follow the task instructions. Participants were psychology undergraduates from Western Sydney University and were reimbursed with credit towards their course. All included participants signed informed consent. None reported any hearing or vision impairments, nor any learning or language disorders.

All participants completed a demographics questionnaire modelled after the Language Experience and Proficiency Questionnaire (LEAP-Q, [18]). The AusE monolinguals reported knowledge of only English and were all born in Australia, except for one participant who moved to Australia at the age of 4. The Arabic–English bilinguals acquired both languages within the first 10 years of life: four acquired both English and Arabic simultaneously, 10 acquired Arabic first, and five acquired English first. In this group, 10 participants were born in Australia, and three arrived in Australia before the age of 5. In the questionnaire, participants also rated their proficiency in reading, writing, speaking, and listening in each of their language(s) on a 9-point Likert scale, with 1 denoting no proficiency, and 9 native proficiency. The bilinguals reported an average score of 5 or greater across the domains of reading, writing, speaking and listening proficiency in English. In Arabic, they also reported an average score of 5 or greater for speaking and listening only, while reading and writing proficiency varied. However, we decided this was acceptable given that only spoken Arabic was presented in the experiment. The bilinguals identified with various cultures, including Syrian, Egyptian, Lebanese, Palestinian, Jordanian, and Saudi. One bilingual participant also spoke French but reported an average score of below 5 for their reading, writing, speaking, and listening proficiency.

## 2.2. Stimulus materials

Target stimuli were sequences of letters from the English alphabet spoken by a female Australian English speaker. The letters were recorded individually in a sound-attenuated booth at a sample rate of 44.1 kHz (16-bit). The letters were then grouped by cognitive load condition, randomised, and concatenated into a five-letter sequence using Praat [19]. Each letter was presented at an interstimulus interval of 250 ms, and stimuli were normalised to 65 dB sound pressure level (SPL). In the low cognitive load condition, the letters were acoustically dissimilar to one another (H, J, K, Q, Y, Z). In the high cognitive load condition, the letters either shared the same onset phoneme /e/ (F, L, M, N, S, X), or the same offset phoneme /i:/ (B, C, D, G, P, T, V), making them acoustically similar.

For the masker conditions, two female native speakers of each language (Australian English, Arabic, Swedish, and Spanish), for a total of eight talkers, produced 336 sentences from the Syntactically Normal Sentence Test list (SNST; [20]). These sentences are controlled for phrase structure and word frequency, and are semantically anomalous (e.g., "The salt dog caused the shoe"), since semantically meaningful masker speech makes speech-in-speech recognition even harder [3]. Like Australian English, which contains 19 vowels (monophthongs and diphthongs), Swedish contains a large number of vowels at 21 [21]. Conversely, Arabic only contains 8 [22], and Spanish 5 [23]. We first created a single masker track per talker by concatenating all of that talker's sentences in a random order. Using a script from Brouwer [24], the long-term average speech spectra for all masker tracks were normalised in Praat to minimise any spectral attributes which could interfere with the linguistic effects of masking. The masker tracks were then normalised to 70 dB SPL.

To create the final stimuli, each letter sequence was then combined with masker speech from both talkers of a language, excised from each masker track at a random starting point. The masker started 500 ms before the onset of the target letter sequence and ended 500 ms after target offset, in order to emulate the continuous stream of background speech encountered in natural listening environments. The final stimuli were presented at a signal-to-noise ratio (SNR) of -5 dB.

## 2.3. Procedure

The experiment was conducted remotely using E-Prime Go 1.0 [25]. Participants were instructed to sit at a table in a quiet environment and used their own computer and headphones. The

task started with four practice trials, presented in quiet (two trials) or +10dB SNR (two trials). Participants then completed the 80 experiment trials (4 maskers × 20 targets) at -5 dB SNR, with half of all trials per masker condition in the low cognitive load condition, and the other half in the high cognitive load condition. Within the high cognitive load condition, 5 trials contained letters sharing an identical onset phoneme, and 5 shared an identical offset phoneme. All conditions were presented in a mixed, randomised order across the 80 test trials.

In each trial, participants heard a sequence of five English letters presented in masker speech. After a 1 second delay from the stimulus offset, participants were prompted to type the letters they heard in the order that they heard them. After pressing the ENTER key to submit their response, they then proceeded to the next letter sequence.

## 3. Results

Data from one AusE monolingual were excluded from analysis, as this participant's mean accuracy in the Arabic, Spanish, and Swedish masker conditions was 3 standard deviations from the group mean. The analyses below thus include data from 60 AusE monolingual participants.

The dependent variable was participants' letter recognition accuracy per trial, measured as the Levenshtein edit distance, which is the minimum number of insertions, deletions, and/or substitutions required to match two strings (i.e., the response and the presented letter sequence) [26]. A lower Levenshtein distance indicates higher similarity between two strings, and thus a better recognition of the target sequence. For the target sequence HJKQY, a response of HJKQY would have a Levenshtein distance of 0 (no edits required), a response of HJKQM would result in a distance of 1 (substitution of the final letter), while ABCDE would have a distance of 5 (all characters substituted). The Levenshtein distance was calculated for each trial using the *stringdist* package [27] in R Version 4.4.1 [28]. For consistency, only the first 5 letters in each response were considered. Mean Levenshtein distances for the AusE monolingual and Arabic–English bilinguals across masker and cognitive load conditions are shown in Figure 1.

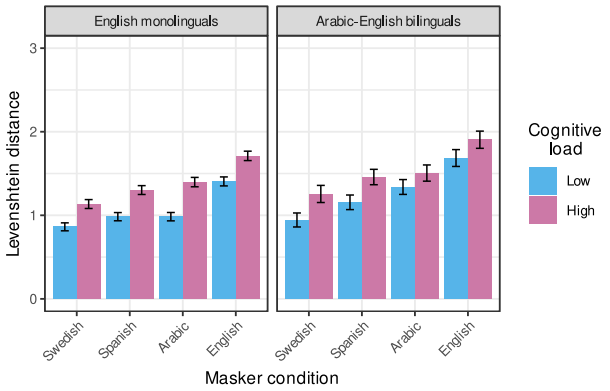


Figure 1: *Mean Levenshtein distance between target sequence and response by listener group across masker and cognitive load. Error bars depict the standard error.*

Using the *brms* package [29] in R, a Bayesian generalised linear mixed model was fitted to the data, with the treatment-coded fixed effects Listener Group (AusE [reference level], Arabic–English), Masker (English [reference level], Arabic, Swedish, Spanish), and Cognitive Load (Low [reference level], High). To determine whether there was a learning effect, the Trial Number variable was standardised to a mean of 0 and standard deviation of 1 using R's *scale* function, then included as a fixed factor. The model also included the random intercept for item and by-participant random slopes for Listener Group. We used weakly informative priors (i.e., a normal distribution with a mean of 0 and a standard deviation of 1), and a binomial distribution.

Directional hypothesis tests were run based on the following predictions. For both listener groups, a higher Levenshtein distance (i.e., poorer letter sequence recognition) would be observed for the English masker condition (identical language to target) compared to the other maskers. For the monolinguals, the Swedish masker would have the second highest Levenshtein distance (phonetically similar to target).

Table 1. *Directional hypothesis tests for the effects of Listener Group, Masker, and Cognitive Load. Mean = mean posterior distribution, 90% CI = one-sided 90% credibility interval, ER = evidence ratio, PP = posterior probability. As the dependent variable is inversely related to performance, a hypothesis of the form* a > b *means that* a *is hypothesised to affect performance more negatively than* b.

| Hypothesis | Mean | 90% CI | ER | PP |
|---|---|---|---|---|
| **AusE group** | | | | |
| Swedish < English | -0.67 | [-0.84, -0.50] | > 7999.00 | 1.00 |
| Spanish < English | -0.48 | [-0.65, -0.31] | > 7999.00 | 1.00 |
| Arabic < English | -0.43 | [-0.60, -0.26] | > 7999.00 | 1.00 |
| Spanish < Swedish | 0.19 | [0.02, 0.37] | 0.03 | 0.03 |
| Arabic < Swedish | 0.24 | [0.07, 0.41] | 0.01 | 0.01 |
| **Arabic–English group** | | | | |
| Swedish < English | -0.76 | [-0.97, -0.56] | > 7999.00 | 1.00 |
| Spanish < English | -0.51 | [-0.72, -0.31] | 7999.00 | 1.00 |
| Arabic < English | -0.39 | [-0.59, -0.19] | 1332.33 | 1.00 |
| Spanish < Swedish | 0.06 | [-0.10, 0.21] | 0.38 | 0.27 |
| Arabic > Swedish | 0.13 | [-0.03, 0.28] | 10.41 | 0.91 |
| Arabic > Spanish | 0.07 | [-0.08, 0.22] | 3.73 | 0.79 |
| **Both groups** | | | | |
| High cog. load > low cog. load | 0.37 | [0.25, 0.48] | > 7999.00 | 1.00 |
| Scaled trial < 0 | 0.02 | [-0.01, 0.04] | 0.13 | 0.12 |
| AusE < Arabic–English | 0.25 | [-0.15, 0.64] | 5.81 | 0.85 |

For the bilinguals, the Arabic masker would have the second highest Levenshtein distance (listeners' native language), followed by the Swedish masker. Finally, for both groups, the high cognitive load condition (more acoustically confusable letter sequences) would result in a higher Levenshtein distance than the low cognitive load condition.

Evidence ratios (ER), which is the ratio of the posterior probability (PP) of the test hypothesis and that of the alternate hypothesis, were used to determine whether there was strong evidence for the tested hypotheses. An ER of over 19 is analogous to a p-value of $< 0.05$ in frequentist statistics, and is considered strong evidence in support of the test hypothesis [30].

Hypothesis test results (see Table 1) show that, for the AusE monolinguals, there is extremely strong evidence that the Swedish, Spanish, and Arabic maskers were less detrimental to performance than the English masker, with PPs of 1.00 and ERs of infinity (*inf*). A PP of 1.00 indicates that 100% of the posterior samples fall on the side of the test hypothesis, while an ER of infinity can be read as greater than $S$ - 1, where $S$ is the number of posterior draws used in the model (i.e., 8000, giving an ER of > 7999). For the Arabic–English bilinguals; the Swedish, Spanish, and Arabic maskers were also less detrimental to performance than English maskers. There was also very strong evidence that the high cognitive load condition resulted in higher Levenshtein distance scores than the low cognitive load condition.

However, there was no evidence that a phonetically similar non-English masker (Swedish) was more detrimental to letter sequence recognition than a dissimilar masker (Arabic and Spanish); this was the case for both AusE monolinguals and Arabic–English bilinguals. For the Arabic–English bilinguals, the Arabic masker also was not more detrimental to letter sequence recognition than Swedish or Spanish. The listener groups also did not differ in performance across masker conditions. Furthermore, there was no evidence for any learning effects over the course of the task.

## 4. Discussion

The present study investigated the effects of listener group, masker language, and cognitive load condition during speech-in-speech recognition, using a letter sequence recall task. As predicted, the findings suggest that letter sequence recognition is more difficult when the target and masker language are the same. This is consistent with findings from previous studies on this topic [3]–[5]. We also found that recognition is poorer when the letters in a sequence are acoustically confusable, in line with past studies that also observed this phenomenon, albeit with written stimuli [17]. It should be noted that data collection is ongoing, so a larger sample size for the Arabic–English bilinguals may uncover additional speech-in-speech recognition patterns regarding listeners' familiarity to the masker speech.

Our predictions regarding the phonetic similarity of the non-English maskers were not borne out. The Swedish masker does not appear to hinder letter sequence recognition more than Spanish or Arabic, even though Swedish is more similar to English both phonetically and rhythmically (both languages are stress-timed [21]). In fact, it appears that the Swedish condition was easier to complete than the other maskers. Possible explanations arise for why the Swedish masker was less detrimental to speech-in-speech recognition performance. Swedish, as a pitch-accented language [21], may sound less speechlike and more "musical" to our participants due to its tonal characteristics, allowing our participants to more easily ignore it while attending to the target speech. On a similar note, it is possibly less commonly-heard than Arabic and Spanish by our participants, who reside in Australia. When participants were asked after the task which languages they had heard, none identified Swedish as one of the maskers, but several were able to identify Spanish and Arabic. Future studies can examine these points further, for instance, by using German (which shares the same rhythmic structure as English and a large vowel inventory) as a masker language. Listeners in Australia are also likely to be more exposed to German as a foreign language than Swedish or other Germanic languages. In addition, a study examining various pitch accented and/or lexical tone masker languages could be an avenue to explore whether tonality in languages contributes to how speechlike a masker is perceived.

We also found no evidence that speech-in-speech recognition is inhibited by native knowledge of a masker (when it is *not* the target language). For the Arabic–English bilinguals, listening to an Arabic masker had no negative impact on their performance, even when compared to their monolingual counterparts. In the literature, reports of the impact of listening to a masker known to the listener are varied; poorer speech-in-speech recognition accuracy is more consistently observed in nonnative listeners [9]. The current findings align with studies which have found that monolinguals and bilinguals perform comparably in speech-in-speech recognition, even when the masker is the other language spoken by the bilinguals [14]. The Arabic–English bilinguals in this study were also early bilinguals, and despite currently being English dominant, reported high levels of Arabic speaking and listening proficiency. Furthermore, it should be noted that the variety of Arabic used for our maskers was Modern Standard Arabic. We selected this variety so that participant recruitment would not be restricted to listeners of a single regional variety of Arabic, although it meant that the masker dialect was not native to any of our bilingual participants. While all bilingual listeners correctly identified Arabic as one of the masker languages upon completion of the experiment, this may nevertheless have affected the degree to which they were hindered by the Arabic masker. Future speech-in-speech recognition research could compare the effects of a listener's familiarity to their native regional variety of Arabic versus the prestige Modern Standard Arabic variety.

In sum, speech-in-speech recognition was most difficult when the target and masker language were identical, as well as when the letter sequence was more acoustically confusable. Neither a phonetically-similar masker (i.e., Swedish) nor a non-English masker known to the bilingual listeners (i.e., Arabic) were shown to negatively impact speech-in-speech recognition in this study. Regardless, there is potential for future research into speech-in-speech recognition focusing on suprasegmental qualities of masker languages, such as pitch, as well as the listener's familiarity to dialectal varieties of a language.

## 5. Acknowledgements

We would like to thank Jade Nguyen, Gabrielle Teklic, Bishoy Boulos, Elena Talevska, Adau Aher, Ella Ward, Jasmine Abdel Wahab, Kristina Petkovich, Afrah Parkar, Mai Linh Tran, and Umair Yakub for their contributions to experimental stimuli development and participant recruitment.

## 6. References


[1] Kidd, G. and Colburn, H. S., "Informational masking in speech recognition", in J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, [Eds.], The Auditory System at the Cocktail Party, 60:75–109, Springer International Publishing, 2017.

[2] Pollack, I., "Auditory informational masking", J. Acoust. Soc. Am., 57:S5–S5, 1975.

[3] Brouwer, S., Van Engen, K. J., Calandruccio, L. and Bradlow, A. R., "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content", J. Acoust. Soc. Am., 131:1449–1464, 2012.

[4] Garcia Lecumberri, M. L. and Cooke, M., "Effect of masker type on native and non-native consonant perception in noise", J. Acoust. Soc. Am., 119:2445–2454, 2006.

[5] Van Engen, K. J. and Bradlow, A. R., "Sentence recognition in native- and foreign-language multi-talker background noise", J. Acoust. Soc. Am., 121:519–526, 2007.

[6] Williams, B. T. and Viswanathan, N., "The effects of target-masker sex mismatch on linguistic release from masking", J. Acoust. Soc. Am., 148:2006–2014, 2020.

[7] Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S. and Bradlow, A. R., "Masking release due to linguistic and phonetic dissimilarity between the target and masker speech", Am. J. Audiol., 22:157–164, 2013.

[8] Brown, V. A. et al., "Revisiting the target-masker linguistic similarity hypothesis," Atten. Percept. Psychophys., 84:1772-1787, 2022.

[9] Garcia Lecumberri, M. L., Cooke, M. and Cutler, A., "Non-native speech perception in adverse conditions: A review", Speech Commun., 52:864–886, 2010.

[10] Cooke, M., Garcia Lecumberri, M. L. and Barker, J., "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception", J. Acoust. Soc. Am., 123:414–427, 2008.

[11] Van Engen, K. J., "Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble". Speech Commun., 52:943–953, 2010.

[12] Mayo, L. H., Florentine, M. and Buus, S., "Age of second-language acquisition and perception of speech in noise", J. Speech Lang. Hear. Res., 40:686–693, 1997.

[13] Krizman, J., Bradlow, A. R., Lam, S. S.-Y. and Kraus, N., "How bilinguals listen in noise: Linguistic and non-linguistic factors", Biling. Lang. Cogn., 20:834–843, 2017.

[14] Calandruccio, L. and Zhou, H., "Increase in speech recognition due to linguistic mismatch between target and masker speech: Monolingual and simultaneous bilingual performance", J. Speech Lang. Hear. Res., 57:1089–1097, 2014.

[15] Mattys, S. L., Davis, M. H., Bradlow, A. R. and Scott, S. K., "Speech recognition in adverse conditions: A review", Lang. Cogn. Process., 27:953–978, 2012.

[16] Baddeley, A., "Working memory: looking back and looking forward," Nat. Rev. Neurosci., 4:829–839, 2003.

[17] Conrad, R. and Hull, A. J., "Information, acoustic confusion and memory span", Br. J. Psychol., 55:429–432, 1964.

[18] Marian, V., Blumenfeld, H. K. and Kaushanskaya, M., "The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals", J. Speech Lang. Hear. Res., 50:940–967, 2007.

[19] Boersma, P. and Weenink, D., "Praat: Doing phonetics by computer", 2022.

[20] Nye, P. W. and Gaitenby, J. H., "The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences", in Haskins Laboratories Status Report on Speech Resolution, SR-37/38, 169–190, 1974.

[21] Riad, T., The Phonology of Swedish, Oxford University Press, 2014.

[22] Watson, J. C. E., The Phonology and Morphology of Arabic, OUP Oxford, 2007.

[23] Green, J. N., "Spanish", in M. Harris and N. Vincent, [Eds.], The Romance Languages, 79–130, Taylor & Francis, 1997.

[24] Brouwer, S., "The role of foreign accent and short-term exposure in speech-in-speech recognition", Atten. Percept. Psychophys., 81:2053–2062, 2019.

[25] Psychology Software Tools, Inc., "E-Prime Go." 2020.

[26] Levenshtein, V. I., "Binary codes capable of correcting deletions, insertions, and reversals", Sov. Phys. Dokl., 10:707–710, 1966.

[27] van der Loo, M. P. J., "The stringdist package for approximate string matching", R J., 6:111–122, 2014.

[28] R Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, 2024.

[29] Bürkner, P.-C., "brms: An R package for Bayesian multilevel models using Stan", J. Stat. Softw., 80:1–28, 2017.

[30] Milne, A. J. and Herff, S. A., "The perceptual relevance of balance, evenness, and entropy in musical rhythms", Cognition, 203:104233, 2020.