The Conversation Continues

The Effect of Lyrics and Music Complexity of Background Music on Spoken-Word Recognition

Scharenborg, Odette; Larson, Martha

DOI
10.21437/Interspeech.2018-1088

Publication date
2018

Document Version
Final published version

Published in
Proceedings of Interspeech 2018

Interspeech 2018
2-6 September 2018, Hyderabad

The Conversation Continues: The Effect of Lyrics and Music Complexity of Background Music on Spoken-Word Recognition

Odette Scharenborg[1,2,3] *and Martha Larson*[1,3,4]

[1] Centre for Language Studies, Radboud University Nijmegen, Netherlands
[2] Donders Institute for Brain, Cognition, & Behavior, Radboud University Nijmegen, Netherlands
[3] Intelligent Systems Department, Delft University of Technology, Netherlands
[4] Institute for Computing and Information Sciences, Radboud University Nijmegen, Netherlands
o.e.scharenborg@tudelft.nl, m.larson@let.ru.nl

Abstract

Background music in social interaction settings can hinder conversation. Yet, little is known of how specific properties of music impact speech processing. This paper addresses this knowledge gap by investigating the effect of the 1) complexity of the background music, and 2) the presence versus absence of sung lyrics on spoken-word recognition in background music. To answer these questions, a word identification experiment was run in which Dutch participants listened to Dutch CVC words embedded in stretches of background music in four conditions: low/high complexity and with lyrics/music-only, and at three SNRs. Music stretches with and without lyrics were sampled from the same song in order to control for factors beyond the complexity of the music and the presence of lyrics. The results showed a clear negative impact of more complex music and the presence of lyrics in background music on spoken-word recognition. The results open a path for future work, and suggest that social spaces (e.g., restaurants, cafés and bars) should make careful choices of music to promote conversation.

Index Terms: spoken-word recognition, background music, social settings

1. Introduction

Music is an important part of the soundscape of social interaction settings. In bars, restaurants, and cafés, music serves to communicate information about the setting [1], thus creating an atmosphere. It also promotes conversational privacy [2]. However, the wrong soundscape choices may cause fatigue by increasing the effort necessary to carry on conversation [3], or even disrupt conversation entirely. This work contributes towards the goal of identifying the properties of background music that optimally allow conversations to continue unhindered in social settings. Despite the large body of work on the effect of the presence of background noise on speech processing (see for a review [4]), the influence of specific properties of music on speech processing is not well understood. Here, we focus on the impact of the complexity of the background music and the presence of lyrics on spoken-word recognition [5], while controlling for other factors.

Previous studies have established that music may interfere with speech processing [6],[7],[8],[9] due to both energetic and informational masking [4],[10],[11],[12]. Energetic masking occurs due to the direct interaction of the background music and the speech signal in the same ear [10],[11]. The severity of the masking effect, and thus the reduction in intelligibility of the speech signal, is dependent on the number of "glimpses" still available to the listener [13]. "Glimpses" are time-frequency regions not masked by the background noise that can be used by the listener for speech recognition. Informational masking is the remaining interference after the effect of energetic masking has been taken into account.

The presence of sung lyrics necessarily causes energetic masking. Additionally, studies that have investigated how lyrics in background music affect cognitive tasks suggest that sung lyrics are also a potential source of informational masking, due to their linguistic content. The impact of lyrical vs. non-lyrical music on foreign language vocabulary learning has been studied by [16]. This work found a short-term effect when the language of the sung lyrics was familiar to the learner. The impact of music on work attention was studied by [19], which recommends that music with sung lyrics should be avoided to avoid impact on worker efficiency.

Given the ongoing neuroscience discussion on neural resources sharing between speech and music processing in the brain, cf., [14],[15], one could possibly expect both musical complexity and lyrics to interfere equally with speech perception. However, given the findings on the impact of speech background noise (e.g., [4]), it is also plausible that lyrics in music pose a unique or larger problem for perception than increasing complexity. This we investigate in this study.

We know of three studies that have investigated the effect of background music on speech processing and included background music with sung lyrics [6],[7],[9]. However, none has specifically investigated the role of sung lyrics. Moreover, these studies included different music pieces in the different conditions. In contrast, we isolate the effects of musical complexity and lyrics. Further, in our music-only and sung-lyrics conditions, we aim to control for other musical factors that have been shown to have an effect on speech recognition or learning, i.e., familiarity with the song(s) [6],[16] and the language of the lyrics [16]. For maximum control, we chose to test only one song. We chose Beyoncé's "Formation", the most-searched-for song on Google in 2016 [17]. As reflected in its reviews, e.g., [18], "Formation" is multidimensional and deserves careful listening. Here, however, we have selected it as background music because it is well known and also because it is possible to find stretches of the song with only a minimal beat (low complexity) and stretches of song that layer instrumentals over that beat (high complexity). Comparable stretches exist with and without sung lyrics, allowing us to control for extraneous sources of variation. Finally, we control for age [6] by testing only younger listeners.

Music training/ability has been shown to have a positive effect on speech-in-noise recognition [20], while hearing

10.21437/Interspeech.2018-1088

problems in noisy conditions (even if the listener has no problems in quiet conditions) might have a detrimental effect. We investigate a possible influence of these two factors by including their self-reported musicality and listening problems in background noise in the statistical analysis.

To investigate these questions, a word identification experiment was set-up. Dutch listeners listened to short, CVC Dutch words embedded in background music. Note that "Formation" is an English-language song, but that English-language background music is typical in Dutch social settings. To minimize the influence of higher-order information such as context, words were presented in isolation. Two listening conditions were created: the sung-lyrics condition (music with sung lyrics) and the music-only (music from the same song without lyrics). We expect a larger detrimental effect of the presence of lyrics in the background music on spoken-word recognition than when there are no lyrics present in the background music due to a potential informational masking effect of the lyrics. Similarly, we also expect to observe effects related to music complexity such that music with higher complexity, i.e., with more pulses between the main beats, has a larger masking effect than music with a lower complexity.

2. Experimental set-up

2.1. Participants

Twenty-five native Dutch listeners (21 females; mean age = 22.6, SD = 2.8) from the Radboud University subject pool participated in the experiment. None of the participants reported a history of language, speech, or hearing problems in quiet listening conditions. All participants had (at least) an upper-intermediate proficiency level in English (which is the English proficiency level at the end of Dutch pre-university high schools). The participants were each paid 5 Euros for their participation.

2.2. Materials

2.2.1. Word stimuli

The stimuli consisted of 144 Dutch CVC words spoken by a native speaker of Dutch, and were taken from an earlier study [20], which investigated the role of word frequency and neighborhood density on native spoken-word recognition. The word frequency and neighborhood density of the 144 words were obtained from [22], and were orthogonally varied (but not further investigated in this study).

2.2.2. Background music

The CVC words were embedded in a short sample of background music. For the sung-lyrics condition, we sampled from an original version of the song. For the music-only condition, we sampled from a high-quality version of the song without lyrics that was highly similar to the original version. Similarity was checked by listening to the tracks and through visual inspection of the spectrograms. Both tracks were obtained from YouTube (original: [23]; music only: [24]).

The structure of the song allows us, as mentioned above, to identify comparable stretches of the song with and without sung lyrics having both high complexity (beat and instruments) and low complexity (beat only). Our procedure for creating the stimuli requires sampling background music from a minute-long segment. Since the naturally occurring segments in the song do not represent one continuous minute, we create minute-long segments by selecting stretches by hand and carefully cutting them at the positive-going zero-crossings using *Praat* [25]. We combined the stretches taking care that no abrupt changes in the music or lyrics would occur. The final one-minute segments were checked for naturalness by listening and visually inspecting the spectrograms.

Figure 1 and 2 provide examples of 4 seconds stretches of the low-complexity and high-complexity conditions. The top two panels show the condition with sung lyrics and the bottom two panels the music-only condition. The figures provide visual evidence that our manual sampling process was successful in ensuring that the overall musical and rhythmic structure is the same within each of the complexity conditions (in other words, across the sung-lyrics/music-only conditions).

We sampled from the minute-long segments to create stimuli that combined spoken words and background music at different SNRs, i.e., SNR +15, +5, and 0 dB. A custom-made *Praat* script was used to select random stretches of the minute-long segments and add these stretches to the words. To ensure that the difference between the sung-lyrics and music-only conditions are not related to the lyrics condition having more energy due to the presence of the singing voice compared to the music-only condition, both the words and the randomly selected stretches of background music were set to (an average of) 65 dB prior to setting the SNR. Each word was preceded by 200 ms of leading background music and followed by 200 ms of trailing background music. A Hamming window was applied to the background music, with a fade in / fade out of 10 ms.

The SNRs were determined on the basis of a pilot study with 12 Dutch participants, none of whom participated in the current study. The SNRs were chosen such that for the easiest SNR, the background music is indeed perceived as being in the background, and at a level often found in coffee bars. The more difficult SNRs were chosen as to reflect a situation that is more to be expected in a pub or disco, as we were also interested in whether we could observe a point where the performance would 'break', i.e., would be severely impaired.

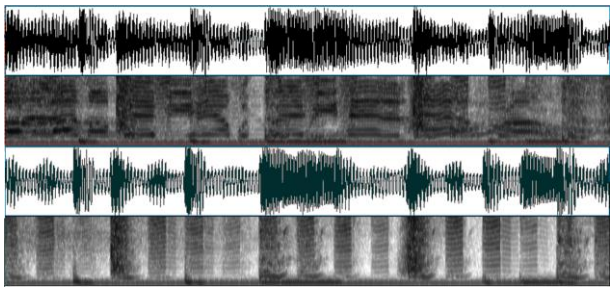

Figure 1. *Waveform and spectrogram of 4 seconds of the low-complexity conditions. Top panels with sung lyrics and bottom panels without lyrics.*

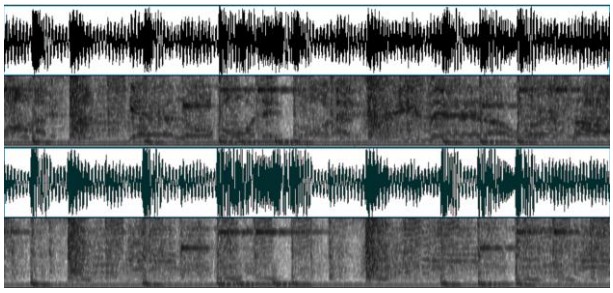

Figure 2. *Waveform and spectrogram of 4 seconds of the high-complexity conditions. Top panels with sung lyrics and bottom panels without lyrics.*

2.3. Procedure

Eight experimental lists were created. Each list consisted of 144 words, with half of the words in the high complexity condition and half of the words in the low complexity condition. Of the low-complexity and high-complexity condition words, half were assigned to the sung-lyrics condition and the other half to the music-only condition, yielding 36 words per complexity/lyrics condition. These 4 sets were each split into three SNR conditions, with 12 words assigned to SNR = 0 dB, 12 words to SNR = 5 dB, and 12 words to SNR = 15 dB. The words were randomly assigned to each of the sets and SNR-conditions. The order of the SNR and sung-lyrics/music-only blocks were randomized and counterbalanced across participants. Each participant was randomly assigned one list.

Participants were tested individually in a sound-treated booth. The stimuli were presented over closed headphones at a comfortable sound level. Participants listened to the 144 words and were asked to type in the word they thought they had heard. After pressing the return key, the next item was played.

After the experiment, listeners filled in a short questionnaire, with questions asking the number of songs they thought were used in the background, whether they were familiar with the song(s), whether they could name the song(s), and whether they played an instrument themselves. Moreover, two questions related to potential hearing problems were asked: one asked whether listeners (were aware of) having hearing problems when listening in quiet (this was used as an exclusion criterion), the other question asked whether listeners experienced problems when listening to speech in background noise, e.g., in a pub.

3. Results

3.1. The effect of complexity and lyrics

Figure 3 shows the proportion of words correctly recognized for each of the SNR conditions for the two complexity conditions and the sung-lyrics and music-only conditions separately. The dashed lines show the results for the low complexity conditions; the solid lines show the results for the high complexity conditions. The solid bullets show the results for the music-only condition; the open bullets show the results for the sung-lyrics condition.

Statistical analyses using generalized linear mixed-effect models (e.g., [26]), containing fixed and random effects, on the accuracy of the recognized words were carried out. The dependent variable was whether the word stimulus was correctly identified ('1') or not ('0'). Fixed factors were SNR (3 levels: +15 dB, +5 and 0 dB (on the intercept)), Lyrics (i.e., the absence (on the intercept) or presence of lyrics in the background music), and the Complexity of the background music (low vs. high (on the intercept)). Stimulus and Subject were entered as random factors. Random by-Subject and by-Stimulus slopes for SNR were added and remained in the best-fitting model.

The results presented here were obtained with the best-fitting model (after model comparisons). This model was obtained by first building the most complex model, i.e., the model with all possible interactions between the predictors. Subsequently, interactions and predictors that proved not significant (at the 5% level) were step-by-step removed from the model, starting with the least significant interaction. The best-fitting model is the model with the lowest AIC.

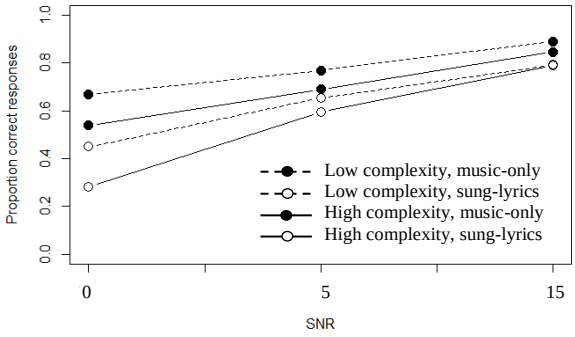

Figure 3: *Proportion of correct responses for the four music background conditions in the three SNR conditions.*

Table 1. *Fixed effect estimates for the best-fitting model for the overall analysis, n=2736.*

Fixed effect	β	*SE*	*p*
Intercept	.223	.270	.410
SNR	.182	.023	< .001
Lyrics	-1.245	.132	< .001
Complexity	.614	.104	< .001
SNR × Lyrics	.036	.018	.052

Table 2. *Fixed effect estimates for the best-fitting model for the analysis with background measures, n=2736.*

Fixed effect	β	*SE*	*p*
Intercept	.673	.296	.023
SNR	.182	.024	< .001
Lyrics	-1.247	.132	< .001
Complexity	.615	.104	.006
Listening Problems	-.854	.308	.004
SNR × Lyrics	.036	.018	.050

Table 1 shows the fixed effect estimates for the best-fitting model of the overall analysis. As expected, significantly more correct answers were given for better SNRs (see effect of SNR in Table 1). Regarding our crucial manipulations, significantly fewer correct answers were given when sung lyrics were present in the background music compared to the music-only condition (Lyrics in Table 1). The marginally significant interaction between SNR and Lyrics indicates that this is more the case in the lower SNR conditions than in the higher SNR conditions (see also Figure 3: the deterioration from SNR +5 to SNR 0 is larger for the sung-lyrics condition compared to the music-only condition). Moreover, significantly more correct answers were given for the low complexity conditions compared to the high complexity conditions.

Analyses of the separate SNR conditions showed that while the significant effect of the presence of lyrics was found at all SNR levels (SNR 0: β=-1.520, SE=.164, p<.001; SNR 5: β=-.821, SE=.160, p<.001; SNR 15: β=-0.880, SE=.228, p<.001), the significant effect of complexity was only found at the two hardest listening conditions (SNR 0: β=.816 , SE=.174 , p<.001; SNR 5: β=.586, SE=.177, p<.001).

3.2. Background questionnaire

The results of the questionnaire showed that most students were not familiar with the song we chose as background music. Only 4 of subjects indicated thinking to have heard 1 song, most reported hearing 2 or 3 different songs (range 1-5 different songs). On a scale of 1 to 4 (not familiar to very familiar), 6

subjects reported “2” (slightly familiar), while 19 subjects indicated “1” (not familiar at all). None could name the song. Eleven participants indicated some musicality (singing and/or playing an instrument). Thirteen participants indicated having (some) problems listening in the presence of noise (note that this might have hearing or attention related origins). Since the song was equally (un)familiar to all participants, we could not investigate the effect of familiarity with the background music on spoken-word recognition. However, musicality and self-reported listening problems in background noise could and were added to the analyses in the previous paragraph as binary factors to investigate the role of musicality and self-reported listening problems in background noise on spoken-word recognition in background music. Table 2 shows the best-fitting model of the analysis including these background measures. In addition to the earlier found main effects and marginally significant interaction between SNR and Lyrics, we found that listeners with self-reported difficulty listening in noisy backgrounds gave significantly fewer correct answers than listeners with no such self-reported difficulty.

4. Discussion and concluding remarks

To our knowledge, this is the first study that systematically investigates the effect of the complexity of and the presence/absence of sung lyrics in background on spoken-word recognition. Our experimental results extend existing knowledge on the effect of different masker types on spoken-word recognition. Importantly, although the experimental conditions do not reflect realistic scenarios in which the sources of music and speech are spatially distinct, aiding the listener in separation [30], they do provide a baseline for the impact of background music on conversation in social settings. We also note that isolated words are more difficult to recognize than words in continuous speech, which have context. For this reason the observed adverse effects may be less noticeable in natural conversational settings.

The key findings are that word recognition is easier in low-complexity than high-complexity background music, and that the music-only condition outperforms the sung-lyrics condition. So, high-complexity background music has a larger masking effect than low-complexity music. Similarly, background music with sung lyrics has a larger masking effect on word recognition than background music without sung lyrics. This effect might be different for listeners with different English proficiency levels, and larger for lyrics in the listener’s native language, since for background noise containing speech, native language has been reported to have larger masking effects than nonnative language [31].

The song has about 120 beats per minute, which amounts to approximately 1-2 beats per stimulus. However, for the high-complexity condition, many intervening notes are present between the main beats, as can also be seen when comparing the spacing of the energy in Figure 1 and Figure 2. Our results on the effect of complexity are in line with findings from [8], who found a larger masking effect for faster tempos. These results suggest that more complex music is a better energetic masker than less complex music, which is as expected as there are more pulses/beats present in high-complexity music that can interfere with the foreground speech. However, the analyses for the individual SNR conditions indicated that this seems to be primarily the case when the SNR is set relatively low. At relatively high SNRs, no difference between high and low complexity music is found. Relating this back to the soundscapes of social interaction settings: in restaurants, where background music is typically not so loud, the complexity of the music will not matter. In bars and cafés where the music is somewhat louder, the owners might take into account to the complexity of the music if they will want conversations to still take place without too much effort.

The effect of the presence of sung lyrics was found for all SNR conditions. The amount of energy was set equal for the sung-lyrics and music-only conditions, meaning that both conditions had similar amounts of energetic masking. The difference between the two conditions thus should primarily be explained by a difference in the amount of informational masking. Note that potentially, the syllable nuclei were aligned with the beats or sung lyrics in one condition and not the other (this could also potentially explain the difference between the low and high complexity conditions). However, our results from a previous, unpublished, experiment investigating the impact of lyrics and complexity showed the same results [29]. Future research will investigate the relationship between the proportion of ‘glimpses’ that are available to the listener [13] and the music complexity to gain insight in the impact of the sub-syllable level distribution patterns of energy resulting from sung lyrics and different music complexities. Thus, similar to speech processing in noisy backgrounds, also for music backgrounds, the presence of linguistic information results in a larger masking effect (e.g., [10]). When designing soundscapes, these results indicate that if the objective is to “let the conversation continue”, it is better to use music without lyrics.

There are many other factors that potentially influence how music affects conversations in social settings. Above, we already mentioned listener familiarity with the language of the lyrics. Other factors are the relative sound power of the singers with respect to the instruments and the age of the listener, cf. [6],[7], as well as, familiarity with the genre, and familiarity with the specific song cf. [6],[7],[32]. Here, we were able to investigate two other factors that might play a role: where musical background of the listener did not have an effect on word recognition in background music, self-reported listening problems in background noise did significantly reduce the number of correct answers. These results add to existing results on the effect of hearing problems on speech processing (e.g., [33],[34]) by extending it to self-reported listening difficulties in noisy listening conditions in younger adults with otherwise normal hearing. Additionally, there are factors that are related to the ability of listeners to separate streams of sounds. The ability to separate streams has been related to speech comprehension [35]. To understand how listeners’ ability to anticipate the rhythm impacts word recognition, we can move, in the future, to longer samples with more than 1-2 main beats.

To conclude, the results suggest that although both music complexity and the presence of sung lyrics play a role in speech processing in background music, the latter interferes more with speech processing than music complexity. Moreover, self-reported listening problems in noisy backgrounds interferes with speech recognition in music backgrounds. The results open a path for future work, and suggest that social spaces (e.g., restaurants, cafés and bars) should make careful choices of music to promote conversation.

5. Acknowledgements

O.S. was sponsored by a Vidi-grant from NWO (grant number: 276-89-003). M.L. was supported in part by EU FP7 project no. 610594 (CrowdRec). The authors would like to thank the student assistants of the lab of O.S. for help in setting up and running the experiments.

6. References

[1] P.M. Lindborg, "A taxonomy of sound sources in restaurants", *Applied Acoustics*, vol. 110, pp. 297-310, 2016.

[2] T. Kato, M. Oka and H. Mori, "Music recommendation system to be effectively difficult to hear the speech noise", *The SICE Annual Conference*, Japan, pp. 2353-2359, 2013.

[3] J.H. Rindel, "Verbal communication and noise in eating establishments", *Applied Acoustics*, vol. 71, pp. 1156-1161, 2010.

[4] M.L.G. Garcia Lecumberri, M. Cooke, and A. Cutler, "Non-native speech perception in adverse conditions: A review", *Speech Communication*, vol. 52, pp.864-886, 2010.

[5] J.M. McQueen, "Speech perception", In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255-275). London: Sage Publications, 2004.

[6] F. Russo and M.K. Pichora-Fuller, "Tune in or tune out: Age-related differences in listening to speech in music", *Ear Hear.*, vol. 29, pp. 746-760, 2008.

[7] D. Başkent, S. van Engelshoven, and J.J. Galvin, "Susceptibility to interference by music and speech maskers in middle-aged adults", *The Journal of the Acoustical Society of Americ*a, vol. 135, EL147, 2014.

[8] S. Ekstrom and E. Borg, "Hearing speech in music", *Noise Health* vol. 13, pp. 277-285, 2011.

[9] K. Gfeller, C. Turner, J. Oleson, S. Kliethermes, and V. Driscoll, "Accuracy of cochlear implant recipients on speech reception in background music", *Ann. Otol. Rhinol. Laryngol.*, vol. 121, pp. 782-791, 2012.

[10] M. Cooke, M.L. Garcia-Lecumberri, and J. Barker, "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception", *Journal of the Acoustical Society of America.*, vol. 123, no. 1, pp. 414-27, 2008.

[11] S. Mattys, J. Brooks, and M. Cooke, "Recognizing speech under a processing load: Dissociating energetic from informational factors", *Cogn. Psych.*, vol. 59, pp. 203-243, 2009.

[12] B.G. Shinn-Cunningham, "Object-based auditory and visual attention", *Trends in Cognitive Sciences*, vol. 12, pp. 182-186, 2008.

[13] M. Cooke, "A glimpsing model of speech perception in noise", *Journal of the Acoustical Society of America*, vol. 119, pp. 1562-1573, 2006.

[14] A.D. Patel, "Language, music, syntax and the brain", *Nature Neuroscience*, vol. 6, pp. 674-681, 2003.

[15] R. Kunert and L.R. Slev,. "A commentary on: "Neural overlap in processing music and speech", *Front. Hum. Neurosci.*, vol. 9, pp. 330, 2015.

[16] A.M.B. de Groot and H.E. Smedinga, "Let the music play! A short-term but no long-term detrimental effect of vocal background music with familiar language lyrics on foreign language vocabulary learning", *Studies in Second Language Acquisition*, vol. 36, no. 4, pp. 681-707, 2014.

[17] G. Kaufman, "Beyonce tops Google's year-end list of top searches", *Billboard*, 14 December 2016.

[18] A. Macpherson, "Beyoncé's Formation review–a rallying cry that couldn't be more timely", *The Guardian*, 8 February 2016.

[19] Y.-N. Shih, R.-H. Huang, and H.-Y. Chiang. "Background music: Effects on attention performance". *Work 42*, pp. 573-578, 2012.

[20] J. Slater, E. Skoea, D.L. Strait, S. O'Connell, E. Thompson, N. Kraus, "Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program", *Behavioural Brain Research*, vol. 291, pp. 244-252, 2015.

[21] F. Hintz and O. Scharenborg, "Effects of frequency and neighborhood density on spoken-word recognition in noise: Evidence from spoken-word identification in Dutch", *Architectures and Mechanisms for Language Processing (AMLaP)*, Bilbao, Spain, 2016.

[22] V. Marian, J. Bartolotti, S. Chabal, and A. Shook, "CLEARPOND: Cross-Linguistic easy-access resource for phonological and orthographic neighborhood densities", *PLoS ONE*, Vol. 7, no. 8, 2012.

[23] https://www.youtube.com/watch?v=7R5olNEjGG4 (accessed January 2018)

[24] https://www.youtube.com/watch?v=mARkecOmkPg (accessed January 2018)

[25] P. Boersma, and D. Weenink, D. "Praat: doing phonetics by computer [Computer program]", 2013. Retrieved from http://www.praat.org/

[26] R.H. Baayen, D.J. Davidson, and D.M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items", *Journal of Memory and Language*, vol. 59, pp. 390-412, 2008.

[27] M.L. Garcia Lecumberri and M. Cooke, "Effect of masker type on native and non-native consonant perception in noise", *Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2445-2454, 2006.

[28] K.J. Van Engen, "Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble", *Speech Communication*, vol. 52, no. 11-12, pp. 943–953, 2010.

[29] O. Scharenborg and M. Larson. Investigating the effect of music and lyrics on spoken-word recognition. arXiv:1803.05058 [cs.SD]

[30] A. Westermann and J.M. Buchholz, "The effect of spatial separation in distance on the intelligibility of speech in rooms", *Journal of the Acoustical Society of America*, Vol. 137, no. 2, pp. 757-67, 2015. doi: 10.1121/1.4906581.

[31] M.L. Garcia Lecumberri and M. Cooke, "Effect of masker type on native and non-native consonant perception in noise", *Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2445-2454, 2006.

[32] N. Perham and H. Currie, "Does listening to preferred music improve reading comprehension performance?" *Applied Cognitive Psychology, Appl. Cognit. Psychol.*," vol. 28, 279-284, 2014.

[33] A. Pittman, K. Vincent, and L. Carter, "Immediate and long-term effects of hearing loss on the speech perception of children," *J Acoust Soc Am.*, vol. 126, no. 3, pp. 1477-1485. doi: 10.1121/1.3177265.

[34] M.K. Pichora-Fuller and G. Singh, "Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation", *Trends Amplif.*, vol. 10, no. 1, pp. 29-59, 2006.

[35] L.-F. Shi and Y. Law, "Masking effects of speech and music: Does the masker's hierarchical structure matter?" *International Journal of Audiology*, 49, pages 296-308, 2010.