AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

2022, Vol. 28, No. 3, 576–588
https://doi.org/10.1037/xap0000368

“Lass Frooby Noo!” the Interference of Song Lyrics and Meaning on Speech Intelligibility

Susanne Brouwer, Noortje Akkermans, Lieke Hendriks, Hanne van Uden, and Veerle Wilms
Centre for Language Studies, Radboud University

This study examined whether song lyrics and their semantic meaning interfere with speech intelligibility. In three experiments, a total of 108 native Dutch participants listened to Dutch target sentences in the presence of three versions of the pop songs *Last Friday Night (T.G.I.F.)* (Experiment 1) or *Hot N Cold* (Experiment 2a and 2b) by singer Katy Perry at different signal-to-noise ratios. The versions consisted of the original English songs, the karaoke versions of the songs without lyrics, and anomalous versions of the songs in the fictional language *Simlish*, which was created for the video game *The Sims*. The songs were played in chronological (Experiments 1 and 2a) or in random order (Experiment 2b). Participants’ task was to type the target sentence they had heard. In all experiments, speech intelligibility was better in nonlyrical (karaoke) than lyrical music (English and Simlish). In addition, listeners performed better in lyrics without semantic meaning (Simlish) than with semantic meaning (English). Finally, speech intelligibility was better when the song in the background was played in chronological rather than in random order. These findings aid in understanding the mechanisms involved during speech-in-music intelligibility.

Public Significance Statement
This study suggests that the presence of song lyrics and their meaning in the background negatively impact speech intelligibility. Additionally, hearing background music in its natural, chronological as opposed to a random order is beneficial for speech intelligibility.

Keywords: speech intelligibility, music, familiarity, informational masking, energetic masking

Music is commonly experienced in people’s everyday lives. It is customary to hear music in social settings such as bars, restaurants, and shops. As speech is often heard in these “cocktail party” environments (Cherry, 1953, see also McDermott, 2009; Myers et al., 2019), it is important to investigate how people separate music from speech. The aim of the present study is to investigate to what extent the presence of song lyrics and their semantic meaning in the background affect speech intelligibility at different signal-to-noise ratios (SNRs). The findings of this study could have important practical consequences (e.g., which music to use in public places), but also contribute theoretically to understanding the mechanisms involved in speech intelligibility under adverse listening conditions.

A large body of work has examined how the presence of multi-talker babble interferes with speech intelligibility (e.g., Brouwer et al., 2012; Brungart, 2001; Calandruccio et al., 2010, 2013; Garcia Lecumberri & Cooke, 2006; Van Engen, 2010; Van Engen & Bradlow, 2007). These multitalker babble tracks contain linguistic information themselves which could interfere with target speech. Pollack (1975) has proposed a distinction between two types of interference: energetic and informational masking (see also Carhart et al., 1969; Darwin, 2008; Kidd et al., 2008). Energetic masking refers to masking at the auditory periphery and is related to the audibility of the target speech. This type of masking produces partial loss of information due to spectral and temporal overlap between the noise and the target speech. In the case of multitalker babble, not only energetic masking of the target speech is involved, but also informational masking plays an important role. Informational masking occurs in the central auditory system and stems from three different yet overlapping causes: incomplete separation between target and masker, limitations in selective attention to the target, and lack of cognitive resources to process the masker that would otherwise be available to process the target (e.g., Mattys et al., 2012; Shinn-Cunningham, 2008).

This article was published Online First July 29, 2021.
Susanne Brouwer https://orcid.org/0000-0003-0321-8227
Noortje Akkermans https://orcid.org/0000-0001-7550-6792
Lieke Hendriks https://orcid.org/0000-0003-2594-5281
Veerle Wilms https://orcid.org/0000-0003-3837-3331
This study was not funded. There are no conflicts of interest, financial, or otherwise. The data that support the findings of this randomized controlled trial study are available upon request.

The materials, data, and analysis script for all experiments can be retrieved from osf.io/n9kaq.

Susanne Brouwer wrote the article. Susanne Brouwer, Noortje Akkermans, Lieke Hendriks, Hanne van Uden, and Veerle Wilms designed the experiment. Noortje Akkermans, Lieke Hendriks, Hanne van Uden and Veerle Wilms collected the data. Susanne Brouwer analyzed the data. Noortje Akkermans, Lieke Hendriks, Hanne van Uden, and Veerle Wilms provided critical revision of the article.

Correspondence concerning this article should be addressed to Susanne Brouwer, Centre for Language Studies, Radboud University, Postbus 9103, 6500 HD Nijmegen, The Netherlands. Email: s.brouwer@let.ru.nl

It is important to understand which conditions provide a release from energetic and/or informational masking during speech-in-speech intelligibility. Previous research has provided accumulating evidence that interference in speech-in-speech intelligibility decreases in conditions in which target-masker similarity is reduced. Speech intelligibility, for example, increases when the masker is spatially distinct from the target (e.g., Freyman et al., 2001) or when the sex of the target talker differs from the masker talker (e.g., Brungart et al., 2001). In addition, Brouwer et al. (2012) demonstrated that listeners found it harder to understand semantically meaningful target speech in semantically meaningful babble compared to anomalous background babble. In contrast, Calandruccio et al. (2018) showed that semantic meaningfulness of the background speech did not increase interference and that performance was sometimes even better for meaningful rather than anomalous background speech. Their materials were carefully controlled for syntactic structure and syllable count. Similarly, Tun et al. (2002) demonstrated that target recall was not more negatively affected by meaningful background speech than randomly ordered background speech.

Together, previous work on speech-in-speech intelligibility research indicates that similarities between target and background speech do not unconditionally change the degree of interference. The question is to what extent informational and energetic masking components are involved in the different studies. Summers and Roberts (2020) distinguished two discernible components of informational masking: acoustic-phonetic interference and linguistic interference. They defined acoustic-phonetic interference as those aspects of informational masking that hinder the extraction or integration of information about speech articulation carried by the time-varying formant-frequency contours, and linguistic interference as those aspects of informational masking that occur after lexical objects have been formed, such as the intrusion of words from an interfering sentence into the percept of the target sentence (p. 1113). In their study, they presented synthetic target sentences in one ear and synthetic speech-like masker sentences in the other. The maskers were intelligible or unintelligible and contained minimal spectro-temporal differences by introducing running asynchronies between formant tracks. Results demonstrated that target sentence intelligibility was 67% when the sentences were presented alone, but performance dropped to 49% with an unintelligible masker and to 41% with an intelligible masker. The authors estimated that two-third of informational masking is caused by acoustic-phonetic interference and the rest by linguistic interference. Related research has shown that acoustic-phonetic interference depends on frequency variation in the masker, particularly the formant-frequency changes in formant transitions (e.g., Roberts & Summers, 2018, 2020).

Besides the role of similarity in separating the target from background speech, familiarity with the two speech streams also seems to have an effect (Brouwer et al., 2012; Cooke et al., 2008; Domingo et al., 2020; Van Engen, 2010). Van Engen (2010), for example, tested native English listeners and native Mandarin listeners on English target sentences in English and Mandarin babble. Both groups experienced a smaller release from masking in (i.e., more interference from) English versus Mandarin (similarity effect), but the nonnative English listeners experienced greater difficulty when Mandarin babble was in the background compared to English babble (familiarity effect). However, Calandruccio and Zhou (2014) examined monolingual and simultaneous bilinguals who showed the same increase in speech intelligibility when the background language was different from the target (similarity effect), despite the fact that the background language was familiar to one group and not to the other.

Taken together, the speech-in-speech intelligibility studies reveal conflicting findings regarding which factors are important for separating target speech from background babble. Some studies show clear similarity effects (Brouwer et al., 2012; Roberts & Summers, 2018, 2020; Summers & Roberts, 2020), whereas others show no such effects (Tun et al., 2002), or even opposite effects (Calandruccio et al., 2018). Furthermore, some studies demonstrate familiarity effects (Van Engen, 2010) while others do not (Calandruccio & Zhou, 2014). A possible explanation for these conflicting speech-in-speech intelligibility findings is that acoustic differences between maskers are not sufficiently controlled for, which makes it difficult to isolate the exact contributions of informational and/or energetic masking. The question arises to what extent informational and/or energetic masking play(s) a role in separating music from speech. In particular, the present study will focus on two components of informational masking, acoustic-phonetic interference and linguistic interference, as suggested by Summers and Roberts (2020) for speech-in-speech intelligibility. Here, these two types of interference will be applied to speech-in-music intelligibility.

Many previous studies have investigated whether music can have an impact on study productivity (see Miller, 2014; Shek & Schubert, 2009, for a review). This type of work often focused on reading comprehension as it is a common component of homework. Some of those studies reported a positive influence of music on reading comprehension which supports the mood-arousal hypothesis. This hypothesis states that when a person listens to enjoyable music, this influences their mood and can subsequently affect their arousal and cognitive performance (e.g., Downing et al., 2004; Lesiuk, 2005). However, other studies have shown that music causes a distraction from what a person is doing and therefore takes away cognitive resources to obtain information from the task at hand (e.g., Anderson & Fuller, 2010; Pool et al., 2003). There seems to be more consensus on how the presence of lyrics impacts comprehension abilities. In particular, it is suggested that meaningful language in the background limits cognitive semantic capacity and therefore inhibits the processing of other semantic tasks (e.g., Herring & Scott, 2018; Oswald et al., 2000). In the present study, song lyrics will be absent or present, and when present, the content of the lyrics will be semantically meaningful or anomalous.

Researchers have only recently shown interest in how music could interfere with speech intelligibility (e.g., Başkent et al., 2014; Ekstrom & Borg, 2011; Eskridge et al., 2012; Gfeller et al., 2012; Russo & Pichora-Fuller, 2008; Scharenborg & Larson, 2018a, 2018b; Shi & Law, 2010). In Başkent et al. (2014), for example, native Dutch listeners were asked to listen to Dutch target sentences in competing speech, noise, and different types of background music (from popular music to classical pieces) and to repeat back what they had heard. The authors found that performance was worse when noise or one of the popular songs was played in the background. However, it is questionable to what extent this study sufficiently controlled their different background types. For example, the target sentences were recorded by a female speaker, whereas the gender of the talker and the singers of the background music varied. As a consequence, listeners could have used gender voice differences as a separation cue in one condition but not in the other

(Brungart et al., 2001). Furthermore, the types of background music were inherently so different from each other (e.g., one had lyrics, the other did not) that it makes it difficult to directly compare them.

To our knowledge, only two studies have investigated whether background music with lyrics interferes more with speech intelligibility than without lyrics[1] (Scharenborg & Larson, 2018a, 2018b). In Scharenborg and Larson (2018a), Dutch listeners performed a word identification task, in which consonant vowel consonant (CVC) words were embedded in background music of different complexities. Their results showed that music with lyrics had a larger masking effect on spoken word intelligibility than music without lyrics, but this effect was independent of music complexity. Note, however, that the three background songs in their study not only differed in rhythm complexity but also in genre and beats per minute, which made it difficult to isolate the effect of music complexity. As a follow-up, Scharenborg and Larson (2018b) conducted a more controlled study with similar listeners and the same task but they only used one song. The structure of the song consisted of comparable stretches with and without lyrics and had both high complexity (beat and instruments) and low complexity (beat only). The results showed a negative influence of more complex music and the presence of lyrics in background music on identifying CVC words.

Russo and Pichora-Fuller (2008) examined whether familiarity with the song lyrics had an influence on word intelligibility in younger and older adults. They found that the younger adults performed better when the background was familiar than unfamiliar music. The researchers argued that younger listeners were able to use the familiarity of the music to their advantage by paying attention to it. This increased attention to the background music aided in generating expectancies which led to better separation between the target speech and the background music. Similar results have been found by Shi and Law (2010) who demonstrated that randomly played music in the background ("scrambled," see Levitin & Menon, 2003) interferes more than music played in chronological order. Mori et al. (2014) also showed that familiarity with the background music could have a facilitative effect on the ability to concentrate.

The aim of the present study was to examine to what extent the presence of song lyrics and their semantic meaning interfere with speech intelligibility. Note that many of the previous studies have focused on the intelligibility of spoken words in music, while real-life conversations often consist of other word structures and/or a combination of words and sentences. The present study therefore used sentences with keywords as targets. In addition, in contrast to prior work, the present study attempted to maximally control the two effects under investigation by using only one song in the background which was sung by the same singer and contained either no lyrics, meaningful lyrics, or anomalous lyrics. More specifically, native Dutch participants were presented with Dutch target sentences embedded in three versions of the American pop song *Last Friday Night (T.G.I.F.)* (Experiment 1) or *Hot N Cold* (Experiment 2a and 2b) by singer Katy Perry at different SNRs. The three versions consisted of the original English songs, a karaoke version of the songs, and anomalous versions of the songs, all sung by Katy Perry. This way, the amount of energetic masking will broadly be the same for these three conditions (English, Karaoke, Simlish). As a result, if differences are observed across the three conditions, they are mostly dependent on changes in informational masking.

The two pop songs used in the experiments are both well-known, popular, and had world-wide chart success. They were first of all chosen because they are familiar to our listener group (Russo & Pichora-Fuller, 2008). Furthermore, both songs have been covered by Katy Perry in *Simlish*, a fictional language spoken in the life simulation video game *The Sims* (Electronic Arts, 2009). In this game, players create virtual people called "*Sims*." Gamers can simulate their daily activities and help direct their moods and satisfy their desires. Simlish is a language made up of gibberish words which cannot be translated so that the dialogs' meaning would be left open to the imagination of the players (comparable to Jabberwocky speech by Lewis Carroll). The phoneme inventory, intonation, pitch, and word length of Simlish are similar to English, so the only difference between these two languages is that the words of Simlish are meaningless. For example, the song *Last Friday Night* in English is called *Lass Frooby Noo* in Simlish (see Appendices A and B for a direct comparison of the lyrics of the two songs in English and Simlish).

By using these three versions of the same songs, we first examined whether speech-in-music intelligibility is influenced by the presence of lyrics. This question gives insight into the role of similarity between the target and the background music, as the versions with lyrics tap into a more similar level of processing (i.e., they cause both energetic and informational masking) than the ones without lyrics (i.e., energetic masking and substantially less informational masking). Following Scharenborg and Larson's (2018a, 2018b) findings, we predicted that participants would perform better on the karaoke than the English and the Simlish version of the songs, as the karaoke version lacks lyrics that could semantically interfere with the target speech. Note that this prediction is in line with the research on the impact of lyrics on comprehension abilities, which has shown that lyrics limit cognitive semantic ability, and as a result, inhibit processing of the target task (e.g., Herring & Scott, 2018; Oswald et al., 2000).

Second, we looked at whether speech intelligibility is affected differently by meaningful (English) compared to anomalous lyrics in the background music (Simlish). In line with Brouwer et al. (2012), who found that anomalous babble is less detrimental for speech intelligibility than meaningful babble, we expect that the English version of the songs interferes more with speech intelligibility than the Simlish version of the songs. Following Summers and Roberts (2020), the English and Simlish conditions are similar in the amount of energetic masking but differ with respect to which informational masking components are involved. More specifically, the Simlish condition primarily contains acoustic-phonetic interference, whereas the English condition consists of both acoustic-phonetic interference and linguistic interference. This supports the prediction that the English condition will interfere more than the Simlish condition.

However, at the same time, it might also be possible that familiarity with the English version aids listeners to generate expectancies (Russo & Pichora-Fuller, 2008) which makes

[1] Previous studies have also examined how lyrics in background music influence cognitive tasks (e.g., de Groot & Smedinga, 2014; Gonzalez & Aiello, 2019; Martin et al., 1988; Rauscher et al., 1993). For example, de Groot and Smedinga (2014) found a short-term effect on foreign language vocabulary learning when the language of the lyrics was familiar to the learner. Moreover, Shih et al. (2012) found that music with lyrics has a detrimental effect on worker efficiency.

intelligibility better, or that meaningful music does not necessarily interfere more than anomalous music (Calandruccio et al., 2018; Tun et al., 2002). Note that the Simlish versions of the songs were not entirely unfamiliar to our listeners, as many of them indicated to be familiar with the melody.

Finally, to further assess the role of familiarity in separating music from speech, we manipulated the presentation order of the background song. In Experiment 2a, the background song was played in chronological order, whereas it was played in random order in Experiment 2b. Following Russo and Pichora-Fuller (2008) and Shi and Law (2010), it was predicted that speech intelligibility would be better when the song was played chronologically (more familiar) than randomly.

Experiment 1

Method

Participants

Thirty six native Dutch listeners (16 males, 20 females) between the ages of 18 and 26 years old participated in Experiment 1. They reported not to have any hearing problems in a questionnaire. After the experiment, most of the participants indicated through a questionnaire to have heard the song *Last Friday Night (T.G.I.F.)* by Katy Perry in the background during the experiment ($N = 31$). However, they indicated not to have recognized Simlish as the language used in the background music during the experimental items ($N = 25$), even though they indicated that they were familiar with the life simulation game *The Sims* ($N = 34$). Participants self-rated their English listening skills to be on average 4.08 ($SD = .60$) on a scale from 1 to 5 (1 = *no knowledge*, 2 = *beginner*, 3 = *average*, 4 = *advanced*, 5 = *near-native*). Eleven participants indicated that they played a musical instrument, but we did not ask for their proficiency level or years of experience. For all experiments, participation was voluntary, and the experimental protocol was approved by the independent ethics committee of the Radboud University [number 2020-1791].

Materials[2]

The target sentences were taken from three lists (1, 7, and 8) from the revised Bamford-Kowal-Bench (BKB; Bamford & Wilson, 1979) Standard Sentence Test. They were translated from English to Dutch by a native Dutch speaker and were checked by another native Dutch speaker. Each list consisted of 16 short, meaningful sentences and each sentence contained 3–4 keywords (e.g., *ZIJ SCHRIJFT naar haar BROER* “SHE WRITES to her BROTHER”), for a total of 50 keywords per list. The 48 sentences in total were recorded by a female native Dutch speaker (the same speaker as used in Brouwer et al., 2012) in a soundproof booth (22,050 Hz, 24 bit).

The target sentences were mixed together with three types of background music in Audacity©: the song *Last Friday Night (T.G.I.F.)* sung by Katy Perry (English), the karaoke version of this song (Karaoke), and the Simlish version of this song (Simlish, also sung by Katy Perry). Simlish is the nonsense language spoken by characters in the life simulation game *The Sims*. Developers of the game aimed for a universally usable language that would leave enough room for the players of the game to use their own imagination (Barnes, 2020; Boland, 2010, pp. 17–22). For example, they purposely designed Simlish to be semantically anomalous such that gamers would not be able to literally translate it into English. Several popular songs were recorded in Simlish to be featured in the game. Even though they may sound similar to the original songs in English, the lyrics do not carry semantic meaning. Each music track was obtained from YouTube (see References) and was equalized to the same root-mean-square level. Moreover, the long-term average speech spectra of the tracks were normalized[3] in Praat (Boersma, 2001), as a means of reducing unequal amounts of energetic masking between conditions. All three tracks were mono tracks. Piloting helped determine that the target sentences were presented at 55 dB Sound Pressure Level (SPL) and the background music at 70 dB SPL, producing an SNR of −15 dB. Note that the target audio setting is lower than the conventional 60–65 dB SPL. The SNRs had to be adjusted because otherwise the maskers would sound uncomfortably loud in combination with the target.

The order of the target lists was kept constant (first list 1, then 7, and finally 8), but the order of the three types of background music was counterbalanced across lists. In addition, participants’ expectancy was maximized by not disrupting the flow of the song in the background across trials. More specifically, the background music accompanying each sentence picked up where the last one left off. The background music was thus played in chronological order. The background music initiated when Katy Perry’s voice was heard in the English condition (i.e., starting from *there’s a stranger in my bed*) and in the Simlish condition (i.e., starting from *zerpa stamby imba bweb*) to make sure that there was a voice present in these conditions. In order to be as consistent as possible, the background music of the Karaoke condition started at the same time as the other two music maskers. The stretches of background music were never repeated across target sentences. This resulted in 16 target sentences for each condition in which the verse was played in the background. Following previous research (e.g., Brouwer et al., 2012), every trial started with 500 ms of the background music only, followed by the target sentence with background music, and after that 500 ms of only background music again. Each participant was exposed to all three conditions (within-subjects design).

Six practice trials preceded the experimental task such that participants could familiarize themselves with the task. These six target sentences were taken from list 20 of the revised BKB (Bamford & Wilson, 1979) and were as the experimental trials played at 55 dB SPL. Participants heard two sentences for each type of background music in the practice session. The SNR was 0 dB for the first two practice trials, −5 dB for the second two, and −10 dB for the final two practice trials.

[2] The materials, data, and analysis script for all experiments can be retrieved from osf.io/n9kaq.

[3] The long-term average speech spectrum (LTAS) script is available upon request. It was made by Chun Liang Chan in Praat (Boersma, 2001), who loosely based it on the script “ltasnoise.praat” by Veenker, van Delft, and Quené (see Quené & van Delft, 2010). The Praat script takes a directory of stimulus files and scales the intensity of the input sound files to the average long-term average spectrum of those sound files. It matches the duration of the longest stimulus (plus any noise padding specified in the arguments to the script) and scales it to match the average intensity of the stimuli. Two methods of spectral averaging are provided: either (a) calculating the LTAS of each file and averaging them or (b) concatenating the stimuli and breaking into equal sized chunks and averaging the LTASs of the chunks. The script also saves the LTAS object into the output directory (along with the noise file). The authors would like to thank Ann Bradlow and Chun Liang Chan for providing the long-term average speech spectra normalization Praat script.

Procedure

Participants were tested individually behind a laptop in a quiet room. They were wearing headphones. They received written instructions on the screen after signing a consent form. The experiment was presented in the open-source program *OpenSesame* (Mathôt et al., 2012). Participants were asked to listen to Dutch sentences spoken by a native Dutch female speaker in the presence of music. They initiated every trial themselves by pressing the space bar on the laptop. After the initiation of a trial, a target sentence accompanied by background music played. Once the fragment had finished playing, participants were asked to type down what sentence they had heard using a keyboard. If they were unable to hear the entire sentence, they were asked to type in as many individual words as they heard. Each fragment had an approximate length of 2 s. Note that during typing, participants heard no target sentences nor background music. The fragments followed each other chronologically in all three conditions. They were allowed to listen to each sentence in background music only once. The total duration of the task was about 20 min.

Data Analysis

A keyword was identified as correct (i.e., a score of 1) if the word was spelled correctly or minor spelling mistakes were made that did not alter the meaning of the word. Obvious spelling errors (e.g., *romijs* for *roomijs* "ice cream") or homophones (e.g., *bank*, meaning "sofa" or "financial institution") were counted as correct. Agreement errors on the (main) verb of the sentences were ignored because of the ambiguous nature of the pronouns *ze* and *zij* in Dutch ("she" (singular) or "they" (plural) in English). Spelling errors that altered the meaning of the word were counted as incorrect (i.e., a score of 0; e.g., *broek* "pants" for *boek* "book").

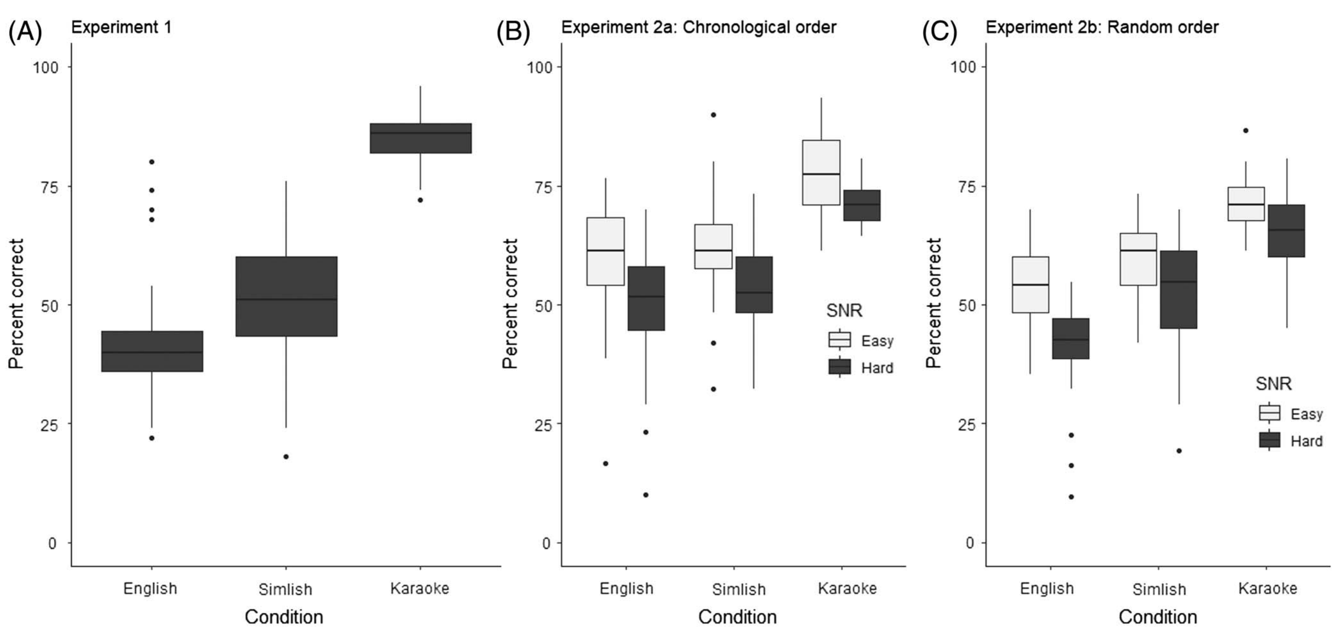
The data were analyzed in R (version 3.2.2; R Core Team, 2018) using the glmer function from the *lme4* package (Bates et al., 2015). Mixed-effects logistic regression analyses were conducted, with keyword identification accuracy as the dichotomous dependent variable (1 = *correct*, 0 = *incorrect*). A logistic linking function was used to deal with the categorical nature of the dependent variable. Condition was entered as fixed effect. Two contrasts were set up. Contrast 1 compared performance between Karaoke versus English and Simlish, thereby testing for the influence of the presence of lyrics on speech-in-music recognition. Contrast 2 compared performance between English and Simlish, thereby testing for the impact of the presence of semantic content on speech-in-music recognition. Participants and items were entered as random effects. Following Barr et al. (2013), we selected the model that maximally converged. The most parsimonious model did not include random slopes.

Results and Discussion

Figure 1A shows the performance of the participants on each type of background music. Recognition accuracy was on average 43.2% (SD = 13.82) in the English condition, 50.0% (SD = 13.83) in the Simlish condition, and 85.2% (SD = 5.56) in the Karaoke condition. The analysis demonstrated an effect of Contrast 1 ($\beta_{\text{CONTRAST1}}$ = 2.74, SE = .10, z value = 27.07, p < .001), indicating that sentence intelligibility was better in music without lyrics (Karaoke) than with lyrics (English and Simlish). In addition, the

Figure 1

Boxplots Showing the Interquartile Ranges of Intelligibility Scores (in % Correct) on Dutch Target Sentence Intelligibility for Each Condition (English, Simlish, Karaoke)



Note. Panel A presents Experiment 1 with one SNR level (−15 dB) for the song "*Last Friday Night*." Panel B and C present Experiment 2a (chronological) and 2b (random) with two SNR levels (easy: −12 dB and hard: −15 dB) for the song "*Hot N Cold*." Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box.

analysis showed an effect of Contrast 2 ($\beta_{CONTRAST2} = -.38$, $SE = .08$, z value $= -4.75$, $p < .001$), revealing that sentences are harder to comprehend in meaningful lyrics (English) than in lyrics without meaningful content (Simlish).

A limitation of these results is that they are based on only one song which was played at one SNR level. There could potentially be something special about this song at this specific SNR level which may have led to these differing degrees of masking, independent of lyrics or meaningful vocals being present. To investigate this, we set up Experiment 2a with a different song at two different SNR levels but the singer remained constant. Note that the data of Experiment 2a have been collected online instead of offline as a consequence of the COVID-19 outbreak.

Experiment 2a

Method

Participants

A total of 56 native Dutch listeners participated in Experiment 2a but 20 of them were not included due to not finishing the experiment ($N = 12$), due to having hearing problems ($N = 6$), or due to not having earphones ($N = 2$). The total sample consisted of 36 participants (12 males, 24 females, age between 18 and 42 year old). In a questionnaire presented after the experiment, they reported to have heard the song *Hot N Cold* by Katy Perry in the background ($N = 34$) without recognizing Simlish as the language used in the background music during the experimental items ($N = 26$), although they indicated that they were familiar with the life simulation game *The Sims* ($N = 29$). They self-rated their English listening skills to be on average 4.22 ($SD = 1.41$) on a scale from 1 (*no knowledge*) to 5 (*near-native*). Nine participants indicated to have played a musical instrument, but we did not ask for their proficiency level or years of music experience.

Materials

The same target sentences were used as in Experiment 1, with the exception of an additional list (List 9) from the revised BKB (Bamford & Wilson, 1979) to have sufficient items for each SNR level. Each list originally consisted of 16 sentences, but we selected 15 sentences of each list (one sentence with four keywords was removed) to have a similar number of sentences and keywords in each condition. Instead of using the song *Last Friday Night (T.G.I.F.)* as in Experiment 1, the song *Hot N Cold* sung by Katy Perry (same singer) was used as background music. The music was played in chronological order as in Experiment 1.

After piloting, the target sentences were presented at 60 dB SPL and the background music at 72 and 75 dB SPL, resulting in two SNRs of −12 dB (Easy) and −15 dB (Hard). The two SNR levels were blocked with the easier condition always presented prior to the harder one. Each block consisted of 30 sentences, 10 sentences per condition. The easy block used Lists 8 and 9, whereas the hard block used Lists 1 and 7, in a fixed order. The conditions were counterbalanced across lists. Every condition at each SNR level consisted of seven target sentences with the verse and three target sentences with the chorus in the background.

Six practice trials from list 20 of the revised BKB (Bamford & Wilson, 1979) played at 60 dB SPL preceded the experimental task. Participants heard two sentences for each type of background music. The SNR was −5 dB for the first two practice trials, −8 dB for the second two, and −10 dB for the final two.

Procedure

The procedure was identical to Experiment 1, except that participants were tested online through the software Qualtrics (Provo, UT). Participants were asked to be in a quiet environment and to wear earphones or headphones. The experiment stopped immediately if participants responded to this questions with the answer "no." They had to actively indicate in a questionnaire whether this was the case. Each trial could be played only once. The task was the same as in Experiment 1. Participants could proceed to the next trial by clicking on an arrow with their mouse. The total duration was about 20 min.

Data Analysis

As in Experiment 1, mixed-effects logistic regression analyses were conducted, with keyword identification accuracy as the dichotomous dependent variable (1 = *correct*, 0 = *incorrect*). Condition and SNR were entered as fixed effects. SNR was numerically contrast coded. For Condition, the same two contrasts as in Experiment 1 were set up. Participants and items were entered as random effects.

Results and Discussion

Figure 1B shows the results of Experiment 2a for each condition at each SNR level. At the easy SNR, recognition accuracy was on average 60.7% ($SD = 12.67$) in the English condition, 62.1% ($SD = 10.92$) in the Simlish condition, and 78.5% ($SD = 9.11$) in the Karaoke condition. At the hard SNR, recognition accuracy was on average 49.1% ($SD = 12.67$) in the English condition, 53.7% ($SD = 9.18$) in the Simlish condition, and 71.1% ($SD = 4.79$) in the Karaoke condition. The results demonstrated a main effect of the first contrast of Condition ($\beta_{CONTRAST1} = 1.58$, $SE = .13$, z value $=$ 12.43, $p < .001$). The analysis showed no effect of the second contrast of Condition ($\beta_{CONTRAST2} = -.08$, $SE = .11$, z value $=$ $-.67$, $p = .50$) and no effect of SNR ($\beta_{SNR} = -1.11$, $SE = .83$, z value $= -1.34$, $p = .18$), but SNR did interact with both condition contrasts. Contrast 1 ($\beta_{SNR:CONTRAST1} = 1.09$, $SE = .22$, z value $=$ 4.87, $p < .001$) revealed that sentence intelligibility increased in music without (Karaoke) compared to with lyrics (English and Simlish), especially at the easier SNR level. Contrast 2 ($\beta_{SNR:CONTRAST2} =$ $-.61$, $SE = .19$, z value $= -3.24$, $p = .001$) showed that sentences are harder to comprehend in meaningful lyrics (English) than lyrics without meaning (Simlish), especially at the harder SNR level.

These data confirm the findings of Experiment 1, indicating that the effect of the presence of lyrics and their meaning on speech intelligibility is robust for both offline and online data collection. As a next step, we replicated Experiment 2a but we presented the song in the background in random instead of chronological order to test for familiarity effects. The random order used here decreased the temporal coherence of the original recordings and therefore disrupted the prosodic and melodic expectancy in all conditions and the syntactic and semantic expectancy in the English condition.

Experiment 2b

Method

Participants

In Experiment 2b, a total of 60 native Dutch listeners participated but 24 of them could not be included due to not finishing the experiment ($N = 18$), hearing problems ($N = 1$), not having earphones ($N = 3$) or being above the age limit ($N = 2$). The final sample consisted of 36 native Dutch listeners (14 males, 22 females, age between 18 and 31 years old). In a questionnaire presented after the experiment, they reported to have heard the song *Hot N Cold* by Katy Perry in the background ($N = 34$) without recognizing Simlish as the language used in the background music ($N = 30$) during the experimental items, although they indicated that they were familiar with the life simulation game *The Sims* ($N = 35$). They self-rated their English listening skills to be on average 4.3 ($SD = .676$) on a scale from 1 (*no knowledge*) to 5 (*near-native*). Six participants indicated to have played a musical instrument, but we did not ask for their proficiency levels or years of music experience.

Materials, Procedure, and Data Analysis

The same materials and online procedure were used as in Experiment 2a, except that the background music was not played in chronological but in random order. For every item, a pseudorandom segment of the background music was combined with a target sentence. This segment had the exact same length as a target sentence but we added 500 ms of additional background music preceding and following each target sentence. Note that these background segments could include words and/or phonemes that were cut halfway at the ends as the length of the segments depended on the length of the target sentence. Importantly, the background music was chronologically played *within* a target sentence but not *between* target sentences. This scrambling method mainly reduced higher level linguistic information, but at the lexical level, individual words remained intelligible and therefore still placed a cognitive load on the semantic system. In other words, the random order decreased the temporal coherence of the original recording and therefore disrupted the prosodic, the syntactic, and the semantic (in the English condition) expectancy across trials. Target sentences were played in the exact same order as in Experiment 2a. Each condition at each SNR level included the same number of target sentences with verse ($N = 7$) or chorus ($N = 3$) in the background as in Experiment 2a. The data analysis was identical to Experiment 2a.

Results

Figure 1C shows the results of Experiment 2b for each condition at each SNR level. At the easy SNR, recognition accuracy was on average 53.5% ($SD = 8.29$) in the English condition, 59.6% ($SD = 8.15$) in the Simlish condition, and 71.7% ($SD = 5.61$) in the Karaoke condition. At the hard SNR, recognition accuracy was on average 41.5% ($SD = 9.69$) in the English condition, 51.9% ($SD = 11.67$) in the Simlish condition, and 65.5% ($SD = 9.60$) in the Karaoke condition. The results demonstrated a main effect of both contrasts of Condition. Contrast 1 showed an increase in recognition accuracy for music without (Karaoke) compared to with lyrics (English and Simlish condition; $\beta_{CONTRAST1} = 1.73$, $SE = .09$, z value $= 18.21$, $p < .001$), whereas Contrast 2 demonstrated better recognition accuracy for anomalous (Simlish) than meaningful lyrics (English; $\beta_{CONTRAST2} = -.75$, $SE = .09$, z value $= -8.28$, $p < .001$). No other effects were significant (all $p > .1$).

Comparing the results of Experiment 2a (chronological) with 2b (random) revealed main effects for both contrast conditions ($\beta_{CONTRAST1} = 1.04$, $SE = .05$, z value $= 22.38$, $p < .001$; $\beta_{CONTRAST2} = -.33$, $SE = .05$, z value $= -6.50$, $p < .001$). It also demonstrated a main effect for Experiment ($\beta_{EXPERIMENT} = -.31$, $SE = .08$, z value $= -3.70$, $p < .001$), indicating that overall performance was better for Experiment 2a than 2b. Most importantly, the effects of Condition and Experiment interacted with each other, indicating that (a) the release from masking between conditions with (English and Simlish) and without lyrics (Karaoke) was larger in Experiment 2a than in 2b ($\beta_{CONTRAST1:EXPERIMENT} = -.18$, $SE = .09$, z value $= 2.03$, $p = .04$), and that (b) the release from masking between conditions with (English) and without meaningful lyrics (Simlish) was larger in Experiment 2b than in 2a ($\beta_{CONTRAST2:EXPERIMENT} = -.26$, $SE = .10$, z value $= -2.68$, $p < .01$). These findings replicate Experiments 1 and 2a and are further discussed in the General Discussion.

General Discussion

The aim of the present study was to examine to what extent song lyrics and their meaning interfere with speech intelligibility. Native Dutch listeners were tested on Dutch target sentences embedded in three different versions of background music at different SNR levels. Across three experiments, two familiar pop songs sung by Katy Perry were presented in their original English forms, as karaoke versions of these songs, and in Simlish, a fictional language spoken in the video game *The Sims*. Importantly, the English and the Simlish version contained lyrics, whereas the karaoke version did not. Moreover, the lyrics of the English version were meaningful for the listeners, whereas those of the Simlish version were anomalous.

In this study, the effects of the presence of song lyrics and their meaning were isolated by using the same song and singer across conditions in each experiment. As a result, the amount of energetic masking was kept as similar as possible across conditions but the amount of informational masking differed across conditions. In particular, the karaoke condition produced less informational masking, whereas the Simlish condition consisted mostly of acoustic-phonetic interference and the English condition of a combination of acoustic-phonetic and linguistic interference (cf. Summers & Roberts, 2020). In addition, the role of familiarity with the background music was investigated by presenting the songs in chronological or in random order.

The results showed three main findings. First, listeners performed better when they heard the songs without (karaoke) than with lyrics (English and Simlish) in the background. This has been shown for two different songs across all experiments in both an offline and an online setting. This reveals that (additional) language-based masking (English and Simlish) has an even more negative influence on speech-in-music intelligibility than energetic masking alone (karaoke). This result was as expected and replicates previous work by Scharenborg and Larson (2018a, 2018b) who also found that music with lyrics has a greater masking effect than music without lyrics. Note however that Scharenborg and Larson (2018b) demonstrated that the effect of lyrics was present at the easier and harder SNR levels, while in the present study (Experiment 2a) it was shown that the effect of lyrics was strongest at the easy SNR level. As different SNR levels have been used across studies it is hard to compare the findings of the two studies and to explain this discrepancy.

The finding that a song without as opposed to with lyrics in the background is beneficial for speech intelligibility, is also in line with previous speech-in-speech intelligibility research in which it is argued that the more similar the target and the background speech are, the harder it is to separate the two (e.g., Brouwer et al., 2012). Furthermore, it is consistent with the work on the impact of music on reading comprehension which has demonstrated that the presence of lyrics limits the use of semantic capacities, and as a consequence, inhibits processing of the target task (e.g., Herring & Scott, 2018; Oswald et al., 2000).

The present study extends previous research by showing that the effect of the presence of lyrics can not only be found for word intelligibility but also for sentence intelligibility. A direct within-subjects comparison between word and sentence intelligibility in background music would be necessary to discover whether the effect of the presence of lyrics is of equal strength across these different tasks. Furthermore, we extended (Scharenborg & Larson, 2018a) and replicated previous research (Scharenborg & Larson, 2018b) by demonstrating that the lyrics effect is present for familiar songs. It would be interesting for future work to directly compare performance on a familiar, an unfamiliar, and their corresponding nonlyrical versions.

The second key finding is that meaningful lyrics (English) interfere more with speech intelligibility than anomalous lyrics (Simlish). This effect of meaning has been demonstrated here for two different songs across all experiments in both an online and an offline environment and is a novel finding within the speech-in-music intelligibility literature. It reveals that the two components of language-based masking, as suggested by Summers and Roberts (2020), indeed show differential effects on speech intelligibility in music. In particular, the English condition consisted of both acoustic-phonetic and linguistic interference, whereas the Simlish condition primarily contained acoustic-phonetic interference. This is in line with the types of errors that were made by the participants. In their responses, they sometimes identified English words in the English condition only. However, they never reported only English lyrics in their responses as the task was to identify words or sentences in a different language (i.e., Dutch). Note also that the most common error made in the English condition was the omission of target words or filling out an incorrect Dutch target word.

The effect of meaning (or linguistic interference) appeared to be strongest at the harder SNR levels when the music was played in chronological order (Experiment 2a). The result is in line with previous speech-in-speech intelligibility work in which it was found that anomalous babble is less detrimental for speech intelligibility than meaningful babble, especially at harder SNR levels (Brouwer et al., 2012). However, this finding goes against prior work by Russo and Pichora-Fuller (2008) who showed that familiarity with the background song might aid listeners in generating expectancies. Following that result, it was expected that meaningful lyrics do not necessarily interfere more than the anomalous lyrics but could even improve speech from music separation. Note, however, that only the lyrics were unfamiliar in the Simlish condition but that the melody of the song was familiar to our listeners. It is therefore possible that the familiarity effect, as described by Russo and Pichora-Fuller (2008), primarily arises when both the lyrics and the melody are known. It would be interesting for future research to follow up on this.

A related possible explanation for the second main finding is that additional resources were required for the English version of the song to suppress the meaningful information. The speech recognition system may have a tendency to remain attuned to meaningful background as a potential source of communicatively relevant information. The Simlish condition, in contrast, could be considered to be less resource demanding, as there was no meaningful information present that needed to be suppressed in order to understand the target sentences. The Simlish condition does not involve suppression of semantic processing and this could therefore have caused participants to perform better in this condition. A similar line of reasoning could also hold for the effect of the presence of lyrics. That is, it may be harder to suppress the lyrical than the nonlyrical information due to the additional communicative information that needs attention. This type of explanation has been given previously for findings in the speech-in-speech recognition literature (e.g., Brouwer, 2019; Brouwer et al., 2012; Summers & Roberts, 2020). Similarly, it also stood out that the target sentences presented in the chorus were more difficult to identify than sentences in verse. This could be due to the fact that the chorus is more complex (Başkent et al., 2014; Scharenborg & Larson, 2018a, 2018b) and/or more familiar (Brouwer et al., 2012; Calandruccio et al., 2010; Garcia Lecumberri & Cooke, 2006; Van Engen, 2010; Van Engen & Bradlow, 2007) and can thus cause a greater informational masking effect. Future research should take this into account by equally dividing the chorus and verse across the sentences or by testing this more directly in two separate conditions.

The third main finding is that speech intelligibility is better when the background music is played in chronological (Experiment 2a) than in random order (Experiment 2b). This result is consistent with previous research (Russo & Pichora-Fuller, 2008; Shi & Law, 2010). Shi and Law (2010) argued that a random order increases stimulus uncertainty, which results in harder listening conditions. Moreover, a random order consists of a reduction in information compared to a chronological one. The current scrambling method mainly reduced higher level linguistic information. At the lexical level, however, individual words remained intelligible and therefore still placed a cognitive load on the semantic system. In other words, the random order decreased the temporal coherence of the original recording and therefore disrupted the prosodic, the syntactic, and the semantic (in the English condition) expectancy across trials. Russo and Pichora-Fuller (2008) state that familiar music, as in the chronological experiment, may be more attended to than to less familiar music, as in the random experiment. Due to the knowledge of the course of the music, listeners could selectively filter information from the target signal at strategic moments. This may lead to more accurate expectancies which could help with target from music separation. The processing demands imposed by familiar music may therefore be reduced compared to less familiar music.

The order effect turned out to have a different influence on the lyrics and the meaning effect. More specifically, the effect of the presence of lyrics was larger for the chronological experiment (Experiment 2a) than the random experiment (Experiment 2b).

It seems that this was primarily due to the high performance on the karaoke condition in the chronological experiment ($M = 74.8\%$, $SD = 8.13$) as opposed to the random experiment ($M = 68.6\%$, $SD = 8.42$) which confirms that a familiar and successively unbroken melody improves speech intelligibility. However, the effect of meaning was larger in the random experiment (Experiment 2b) than the chronological experiment (Experiment 2a). This suggests that meaningful lyrics are especially distracting if the melody of the song is interrupted. In particular, in the chronological experiment the release from masking between the meaningful ($M = 54.9\%$, $SD = 13.98$) and the anomalous condition ($M = 57.9\%$, $SD = 10.87$) is smaller than in the random experiment ($M = 47.5\%$ and $SD = 10.82$ for meaningful; $M = 55.8\%$ and $SD = 10.71$ for anomalous). It is possible that the meaning of the lyrics is less noticeable and/or attracts less attention when the music is played chronologically, as listeners can use the prosody of the song as a cue for speech-in-music separation. However, the meaning of the lyrics may become more important to generate expectations when the music is played at random because in this situation listeners could rely less on the prosody of the song (cf. Russo & Pichora-Fuller, 2008).

There are a number of limitations to the present study of which some could be addressed in future research. First, the present study tested native Dutch participants with English songs in the background which means that the English and Simlish conditions involved nonnative language processing. The reason for choosing English songs rather than Dutch songs is because there are no Dutch songs that are available in Simlish. To get more insight into the effect of native versus nonnative language processing, a follow-up idea would be to test native English participants on English target sentences in the same background songs as the present study or to test native Dutch participants on Dutch target sentences in Dutch background songs.

Second, as melody and lyrics are tightly linked, it is possible that participants "internally heard" the words of a song while listening to a familiar melody. This could be a possible confound in the karaoke condition, as it also started at the time when the lyrics would begin in the song. This might have activated semantic content (i.e., linguistic masking) for the listener that we were not able to control for. Comparing the effect of familiar with unfamiliar (karaoke) songs on speech intelligibility would give more insight into this issue.

A third possible limitation of this study could be that the three tracks used in Experiment 1 and the three tracks used in Experiment 2a and 2b were not completely identical. Although the only difference between the tracks is the presence of lyrics, we did not use one instrumental version of the song to later map the English and Simlish lyrics onto it, but rather downloaded six different versions of the songs from YouTube. There could therefore have been slight differences between the instrumental properties of the tracks. However, based on inspection of the spectral properties of the different tracks we do not expect this to have a major impact on the results.

In conclusion, background music is an extremely common component in our daily lives which makes it important to understand its exact influence on conversations. The aim of the present study was to examine to what extent the presence of song lyrics and their meaning impact speech intelligibility. It was found that both have a negative influence on speech intelligibility. In addition, hearing music in its natural, chronological as opposed to a random order aids intelligibility.

References

Anderson, S. A., & Fuller, G. B. (2010). Effect of music on reading comprehension of junior high school students. *School Psychology Quarterly*, *25*(3), 178–187. https://doi.org/10.1037/a0021213

Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (Eds.), *Speech–hearing tests and the spoken language of hearing-impaired children* (pp. 148–187). Academic Press.

Barnes, A. (2020, February 4). *The Sims turns 20: Creator will wright reflects on the battle he waged to get one of the best games of all time made*. https://www.gamesradar.com/the-making-of-the-sims/

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Başkent, D., van Engelshoven, S., & Galvin, J. J., III (2014). Susceptibility to interference by music and speech maskers in middle-aged adults. *The Journal of the Acoustical Society of America*, *135*(3), EL147–EL153. https://doi.org/10.1121/1.4865261

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Boland, E. (2010). *The sims: The complete guide*. WTYW7.

Brouwer, S. (2019). The role of foreign accent and short-term exposure in speech-in-speech recognition. *Attention, Perception & Psychophysics*, *81*, 2053–2062. https://doi.org/10.3758/s13414-019-01767-8

Brouwer, S., Akkermans, N., Hendriks, L., van Uden, H., & Wilms, V. (2021, March 3). *"Lass Frooby Noo!" The interference of song lyrics and meaning on speech intelligibility*. osf.io/n9kaq

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3),1101–1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5 Pt. 1), 2527–2538. https://doi.org/10.1121/1.1408946

Calandruccio, L., Brouwer, S., Van Engen, K. J., Bradlow, A. R., & Dhar, S. (2013). Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *The American Journal of Audiology*, *22*(1), 157–164. https://doi.org/10.1044/1059-0889(2013/12-0072)

Calandruccio, L., Buss, E., Bencheck, P., & Jett, B. (2018). Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition? *The Journal of the Acoustical Society of America*, *144*(6), 3289–3302. https://doi.org/10.1121/1.5081679

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, *128*(2), 860–869. https://doi.org/10.1121/1.3458857

Calandruccio, L., & Zhou, H. (2014). Increase in speech recognition due to linguistic mismatch between target and masker speech: Monolingual and simultaneous bilingual performance. *Journal of Speech, Language, and Hearing Research*, *57*(3), 1089–1097. https://doi.org/10.1044/2013_JSLHR-H-12-0378

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694–703. https://doi.org/10.1121/1.1911445

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 1262–2527. https://doi.org/10.1121/1.1907229

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414–427. https://doi.org/10.1121/1.2804952

Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1011–1021. https://doi.org/10.1098/rstb.2007.2156

de Groot, A. M. B., & Smedinga, H. E. (2014). 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*, *36*(4), 681–707. https://doi.org/10.1017/S0272263114000059

Domingo, Y., Holmes, E., & Johnsrude, I. S. (2020). The benefit to speech intelligibility of hearing a familiar voice. *Journal of Experimental Psychology: Applied*, *26*(2), 236–247. https://doi.org/10.1037/xap0000247

Downing, J. A., Carlson, J. K., Hoffman, J., Gray, D., & Thompson, A. (2004). A musical interlude using music and relaxation to improve reading performance. *Intervention in School and Clinic*, *39*(4), 246–250. https://doi.org/10.1177/10534512040390040801

Ekstrom, S., & Borg, E. (2011). Hearing speech in music. *Noise & Health*, *13*(53), 277–285. https://doi.org/10.4103/1463-1741.82960

Electronic Arts. (2009). *The Sims 3* [Video Game].

Eskridge, E. N., Galvin, J. J., III, Aronoff, J. M., Li, T., & Fu, Q. J. (2012). Speech perception with music maskers by cochlear implant users and normal-hearing listeners. *Journal of Speech, Language, and Hearing Research*, *55*(3), 800–810. https://doi.org/10.1044/1092-4388(2011/11-0124)

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5 Pt. 1), 2112–2122. https://doi.org/10.1121/1.1354984

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–2454. https://doi.org/10.1121/1.2180210

Gfeller, K., Turner, C., Oleson, J., Kliethermes, S., & Driscoll, V. (2012). Accuracy of cochlear implant recipients in speech reception in the presence of background music. *The Annals of Otology, Rhinology, and Laryngology*, *121*(12), 782–791. https://doi.org/10.1177/000348941212101203

Gonzalez, M. F., & Aiello, J. R. (2019). More than meets the ear: Investigating how music affects cognitive task performance. *Journal of Experimental Psychology: Applied*, *25*(3), 431–444. https://doi.org/10.1037/xap0000202

Herring, D. & Scott, J. (2018). The effect of lyrical and instrumental music on reading comprehension tasks. *Journal of Emerging Investigators*, 1–6. https://emerginginvestigators.org/articles/the-effect-of-lyrical-and-instrumental-music-on-reading-comprehension-tasks

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. Yost (Ed.), *Springer handbook of auditory research 29: Auditory perception of sound sources* (pp. 143–190). Springer. https://doi.org/10.1007/978-0-387-71305-2_6

Lesiuk, T. (2005). The effect of music listening on work performance. *Psychology of Music*, *33*(2), 173–191. https://doi.org/10.1177/0305735605050650

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, *20*(4), 2142–2152. https://doi.org/10.1016/j.neuroimage.2003.08.016

Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*(4), 382–398. https://doi.org/10.1016/0749-596X(88)90063-0

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McDermott, J. H. (2009). The cocktail party problem. *Current Biology*, *19*(22), R1024–R1027. https://doi.org/10.1016/j.cub.2009.09.005

Miller, C. (2014). The differentiated effects of lyrical and non-lyrical music on reading comprehension. *Theses and dissertations*. Rowan University.

Mori, F., Naghsh, F. A., & Tezuka, T. (2014). *The effect of music on the level of mental concentration and its temporal change* [Conference session]. Proceedings of the 6th International Conference on Computer Supported Education (pp. 34–42), Barcelona, Spain. https://doi.org/10.5220/0004791100340042

Myers, B. R., Lense, M. D., & Gordon, R. L. (2019). Pushing the envelope: Developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain Sciences*, *9*(3), Article 70. https://doi.org/10.3390/brainsci9030070

Oswald, C. J., Tremblay, S., & Jones, D. M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, *8*(5), 345–350. https://doi.org/10.1080/09658210050117762

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5. https://doi.org/10.1121/1.1995329

Pool, M. M., Koolstra, C. M., & Voort, T. H. (2003). The impact of background radio and television on high school students' homework performance. *Journal of Communication*, *53*(1), 74–87. https://doi.org/10.1111/j.1460-2466.2003.tb03006.x

Quené, H., & van Delft, L. (2010). Non-Native durational patterns decrease speech intelligibility. *Speech Communication*, *52*(11–12), 911–918. https://doi.org/10.1016/j.specom.2010.03.005

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, *365*(6447), Article 611. https://doi.org/10.1038/365611a0

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Roberts, B., & Summers, R. J. (2018). Informational masking of speech by time-varying competitors: Effects of frequency region and number of interfering formants. *The Journal of the Acoustical Society of America*, *143*(2), 891–900. https://doi.org/10.1121/1.5023476

Roberts, B., & Summers, R. J. (2020). Informational masking of speech depends on masker spectro-temporal variation but not on its coherence. *The Journal of the Acoustical Society of America*, *148*(4), 2416–2428. https://doi.org/10.1121/10.0002359

Russo, F. A., & Pichora-Fuller, M. K. (2008). Tune in or tune out: Age-related differences in listening to speech in music. *Ear and Hearing*, *29*(5), 746–760. https://doi.org/10.1097/AUD.0b013e31817bdd1f

Scharenborg, O., & Larson, M. (2018a). Investigating the effect of music and lyrics on spoken-word recognition. arXiv:1803.05058 [cs.SD]

Scharenborg, O., & Larson, M. (2018b). The conversation continues. The effect of lyrics and music complexity of background music on spoken-word. In B. Yegnanarayana (Ed.), *Proceedings of interspeech 2018* (pp. 2280–2284). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2018-1088

Shek, V., & Schubert, E. (2009, December 3–4). *Background music at work—a literature review and some hypotheses* [Conference session]. The Second International Conference on Music Communication Science, Sydney, Australia.

Shi, L.-F., & Law, Y. (2010). Masking effects of speech and music: Does the masker's hierarchical structure matter? *International Journal of Audiology*, *49*(4), 296–308. https://doi.org/10.3109/14992020903350188

Shih, Y.-N., Huang, R.-H., & Chiang, H.-Y. (2012). Background music: Effects on attention performance. *Work (Reading, Mass.)*, *42*(4), 573–578. https://doi.org/10.3233/WOR-2012-1410

Shinn-Cunningham, B. G. (2008). Object-Based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Summers, R. J., & Roberts, B. (2020). Informational masking of speech by acoustically similar intelligible and unintelligible interferers. *The Journal of the Acoustical Society of America*, *147*(2), 1113–1125. https://doi.org/10.1121/10.0000688

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*(3), 453–467. https://doi.org/10.1037/0882-7974.17.3.453

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11–12), 943–953. https://doi.org/10.1016/j.specom.2010.05.002

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519–526. https://doi.org/10.1121/1.2400666

YouTube. (2011). *English version of the song Hot N Cold by Katy Perry*. https://www.youtube.com/watch?v=lz0JazR4-sI

YouTube. (2011). *English version of the song Last Friday Night (T.G.I.F.) by Katy Perry*. https://www.youtube.com/watch?v=aJwrP501Msw

YouTube. (2012). *Karaoke version of the song Hot N Cold by Katy Perry*. https://www.youtube.com/watch?v=rRVdEQxqNKs

YouTube. (2012). *Karaoke version of the song Last Friday Night (T.G.I.F.) by Katy Perry*. https://www.youtube.com/watch?v=AORNoOZJDp4

YouTube. (2012). *Simlish version of the song Hot N Cold by Katy Perry*. https://www.youtube.com/watch?v=ucYv1zX13zU&t=164s

YouTube. (2012). *Simlish version of the song Last Friday Night (T.G.I.F.) by Katy Perry*. https://www.youtube.com/watch?v=NkdrJWE-rqI

Appendix A

Lyrics of the English and the Simlish Version of the Song Last Friday Night (T.G.I.F.) by Katy Perry

English: Last Friday Night (T.G.I.F.)	Simlish: Lass Frooby Noo
There's a stranger in my bed	Zerpa stamby imba bweb
There's a pounding in my head	Zerpa powey imba heb
Glitter all over the room	Nifler aba reba roo
Pink flamingos in the pool	Fweeka minzo eba foo
I smell like a minibar	Cowsa lovy minza bar
DJ's passed out in the yard	Ze yay blousa iza bar
Barbies on the barbeque	Iba umba derpa cu
This a hickie or a bruise?	Zeesa ika uba broov
Pictures of last night ended up online	Finchy zub lep sny
I'm screwed, oh well	Eby up obly maskoo, oh welb
It's a blacked-out blur, but I'm pretty sure it ruled	Epsa berp ta bur, aza pipty shner zaroo
Damn	Dwam
Last Friday night	Lass frooby noo!
Yeah, we danced on tabletops	Yarby dansel dabel doops
And we took too many shots	Imi dooka mimi shoops
Think we kissed, but I forgot	Sipi gibsy a fergoob
Last Friday night	Lass frooby noo!
Yeah, we maxed our credit cards	Yarby meksa crabit car
And got kicked out of the bar, so we hit the boulevard	Inna keet it towy yar
	So be hipta bu leeyar
Last Friday night	Lass frooby noo!
We went streaking in the park	Ima stika ina par
Skinny dipping in the dark, then had a ménage à trois	Skeeby deeby mina yar
	Dina hana showa tar
Last Friday night	Lass frooby noo!
Yeah, I think we broke the law	Yarma tinka bookey yow
Always say we're going to stop-op, oh woah	Owa sina go estow, ow oh-oh
But this Friday night, do it all again	Badipsa frooby noo, dukwey ahhh da kweeb
But this Friday night, do it all again	Badipsa frooby noo, dukwey ahhh da kweeb
Trying to connect the dots	Topy nu conecsa dops
Don't know what to tell my boss	Duna waka dena bops
Think the city towed my car	Tinka siby duna ka
Chandelier is on the floor	Tomba looey yisa fla
Ripped my favorite party dress	Rempy ferva perba dets
Warrant's out for my arrest	Wara oofa mona reks
Think I need a ginger ale	Tinka neeba jamberay
That was such an epic fail	Towa sooshka neeba fay
Pictures of last night ended up online	Finchy zub leb sny
I'm screwed, oh well	Eby up obly maskoo, oh welb
It's a blacked-out blur, but I'm pretty sure it ruled	Eska berb ta bur, aza pipty shner zaroo
Damn	Dwam
Last Friday night	Lass frooby noo

(*Appendices continue*)

Appendix A (*continued*)

English: Last Friday Night (T.G.I.F.)	Simlish: Lass Frooby Noo
Yeah, we danced on tabletops	Yarby dansel dable doops
And we took too many shots	Imi dooka mimi shoops
Think we kissed, but I forgot	Sipi gibsy a fergoob
Last Friday night	Lass frooby noo
Yeah, we maxed our credit cards	Yarby meksa crabit car
And got kicked out of the bar	Inna keet it towy yar
So we hit the boulevard	So be hipta bu leeyar
Last Friday night	Lass frooby noo
We went streaking in the park	Ima stinka ina par
Skinny dipping in the dark	Skeeby deeby mina yar,
then had a ménage à trois	dina hana showa tar
Last Friday night	Lass frooby noo
Yeah, I think we broke the law	Yarma tinka bookey yow
Always say we’re going to stop-op, oh woah	Owa sina go estow, ow oh-oh
But this Friday night, do it all again	Badipsa frooby noo, dukwey ahhh da kweb
This Friday night, do it all again	Badipsa frooby noo, dukwey ahhh da kweb
(Do it all again) This Friday night	Badipas frooby noo
T.G.I.F., T.G.I.F., T.G.I.F.	T.G.I.F, T.G.I.F, T.G.I.F.
T.G.I.F., T.G.I.F., T.G.I.F.	T.G.I.F, T.G.I.F, T.G.I.F.
Last Friday night	Lass frooby noo
Yeah, we danced on tabletops	Yarby dansel dable doops
And we took too many shots	Imi dooka mimi shoops
Think we kissed, but I forgot	Sipi gibsy a fergoob
Last Friday night	Lass frooby noo
Yeah, we maxed our credit cards	Yarby meska crabit car
And got kicked out of the bar,	Inna keet it towy yar
so we hit the boulevard	So be hipta bu leeyar
Last Friday night	Lass frooby noo
We went streaking in the park	Ima stika ina par
Skinny dipping in the dark,	Skeeby deeby mina yar,
then had a ménage à trois	dina hana showa tar
Last Friday night	Lass frooby noo
Yeah, I think we broke the law	Yarma tinka bookey yow
Always say we’re going to stop-op, oh woah	Owa skina yo estow, ow oh-oh
But this Friday night, do it all again	Badipsa frooby noo, dukwey ahhh da kweb

Appendix B

Lyrics of the English and the Simlish Version of the Song Hot N Cold by Katy Perry

English: Hot N Cold	Simlish: Ninap cou
You change your mind like a girl changes clothes	Vous chikanip laka gurl chika claps
Yeah you PMS like a bitch, I would know	Va vous zee e rex laka brich, aba nup
And you always think, always speak cryptically	Om vous olweh zonk, ibu seek cryptozeek
I should know that you’re no good for me	Ashanup za vous ku da vina
Cause you’re hot then you’re cold	Kavahap ninap cou
You’re yes then you’re no	Fahit tinap blo
You’re in then you’re out	Vaip tinip aw
You’re up then you’re down	Vanip tinip taw
You’re wrong when it’s right	Va longe rishwai
It’s black and it’s white	Zi hurkinish why
We fight, we break up	Wi fout, we wi dap
We kiss, we make up	Wi kip, we ni daw
(You) you don’t really wanna stay, no	(Vous) vous nu really wanna sti, oh
(You) but you don’t really wanna go, oh	(Vous) vous nu really wanna cri, oh
Cause you’re hot then you’re cold	Kavahap ninap cou
You’re yes then you’re no	Fahit tinap blo
You’re in then you’re out	Vaip tinip aw
You’re up then you’re down	Vanip tunip taw
We used to be just like twins, so in sync	Wa yip tu bey juk ku twip, su ba sip
The same energy now’s a dead battery	Ta si enerjonk nawza dwip batteronk

(*Appendices continue*)

Appendix B (*continued*)

English: Hot N Cold	Simlish: Ninap cou
Used to laugh 'bout nothing, now you're plain boring	Yiptulorp ba nothing, no yoh playy amorengk
I should know that you're not gonna change	Ashanup za u kudi ki ye
Cause you're hot then you're cold	Kavahap ninap cou
You're yes then you're no	Fahit tinap blo
You're in then you're out	Vaip tinip aw
You're up then you're down	Vanip tinip taw
You're wrong when it's right	Va longe rishwai
It's black and it's white	Zi hurkinish why
We fight, we break up	Wi fout, we wi dap
We kiss, we make up	Wi kip, we ni daw
(You) you don't really wanna stay, no	(Vous) vous nu really wanna sti, oh
(You) but you don't really wanna go, oh	(Vous) vous nu really wanna cri, oh
Cause you're hot then you're cold	Kavahap ninap cou
You're yes then you're no	Fahit tinap blo
You're in then you're out	Vaip tinip aw
You're up then you're down	Vanip tunip taw
Someone, call the doctor	Sugnorg, cul di docturg
Got a case of love bipolar	Gutta case o luv bicoler
Stuck on a roller coaster	Stucowa roli nowster
Can't get off this ride	Can't rinup za wey
You change your mind	Vous chikanip
Like a girl changes clothes	Laka gurl chika claps
Cause you're hot then you're cold	Kavahap ninap cou
You're yes then you're no	Fahit tinap blo
You're in then you're out	Vaip tinip aw
You're up then you're down	Vanip tinip taw
You're wrong when it's right	Va longe rishwai
It's black and it's white	Zi hurkinish why
We fight, we break up	Wi fout, we wi dap
We kiss, we make up	Wi kip, we ni daw
You're hot then you're cold	Vahap ninap cou
You're yes then you're no	Fahit tinap blo
You're in then you're out	Vaip tinip aw
You're up then you're down	Vanip tinip taw
You're wrong when it's right	Va longe rishwai
It's black and it's white	Zi hurkinish why
We fight, we break up	Wi fout, we wi dap
We kiss, we make up	Wi kip, we ni daw
(You) you don't really wanna stay, no	(Vous) vous nu really wanna sti, oh
(You) but you don't really wanna go, oh	(Vous) vous nu really wanna cri, oh
Cause you're hot then you're cold	Kavahap ninap cou
You're yes then you're no	Fahit tinap blo
You're in then you're out	Vaip tinip aw
You're up then you're down	Vanip tunip taw

Received October 21, 2020
Revision received April 22, 2021
Accepted April 26, 2021 ■