Hear No Evil: Can Music Attenuate the Irrelevant Speech Effect?

Wei Jie Soh (weijiesoh31@gmail.com)
Stephen Wee Hun Lim (psylimwh@nus.edu.sg)
Department of Psychology, National University of Singapore
9 Arts Link, Singapore 117570

Abstract

This study aims to replicate the *irrelevant speech effect* (ISE) in a local context and, more important, is the first to directly investigate if musical information can reduce impairments imposed by the ISE on a serial word recall task. Thirty-five undergraduates from the National University of Singapore performed serial recall on 10 word lists. The lists were presented under 5 auditory conditions, namely: *Music-Only*, *Combined* (music with background speech), *Scrambled music with background speech, Background Speech-Only* and *White Noise* conditions. The *Scrambled* condition contained the same piece of music as the *Combined* condition except that it was re-arranged in a random fashion; the mission of this condition was to specifically provide a comparison basis to test if "musical structure" *per se* actually attenuates the ISE. A significant main effect of music conditions emerged. ISE was successfully replicated, where a significantly lower percentage of correct words was recalled in the *Background Speech-Only* condition compared to all other conditions. ISE was also successfully attenuated, but the present data suggest that musical structure *per se* was not (at least not entirely) responsible for the attenuation, since the *Scrambled* condition had superior performance than both the *Combined* and *Background Speech-Only* conditions. Here, we propose and discuss several novel theoretical models involving changing acoustical features, selective attention, and arousal to account for the present findings.

Keywords: Irrelevant speech effect; music; recall performance.

Introduction

The *irrelevant speech effect* (ISE) is the finding that background speech significantly impairs serial recall performance, even when the background speech is irrelevant to the task (Farley, Neath, Allbritton & Surprenant, 2007). First demonstrated by Colle and Walsh (1976), the researchers presented subjects with lists of eight consonant-items visually together with a passage read out in German. The background speech was considered irrelevant as participants were told to ignore the passage and that no subsequent recall of the background speech was required. Serial recall was significantly impaired in the irrelevant speech condition compared to the quiet (control) condition. The ISE is found to be robust and independent of speech intensity, within the range of 40 to 76 dB (Ellermeier & Hellbrück, 1998). The effect is also significant regardless of whether the irrelevant speech is presented together with or after the word list (Miles, Jones & Madden, 1991), and evident over repeated trials or sessions (Tremblay & Jones, 1998).

The question of greater interest (and importance) is whether one could ever circumvent the ISE, given the potential costs on cognitive performance that are associated with the negative impacts of ISE under a variety of situations. A possible candidate to abate irrelevant speech is instrumental music, due to the fact that music has been found to modulate work performance. Lesiuk (2010), for instance, found that listening to preferred music led to improvements in performance within the context of highly cognitive demanding jobs.

This study had two goals. The first was to (first) replicate the ISE in a local context (among undergraduates at the National University of Singapore). The background speech, accordingly, comprised of contents related to undergraduates, ranging from modules, bid points, gossip and current news ensuring that the contents were distracting enough while trying to concentrate on learning a word list. Second, and more important, this study aimed to discover whether instrumental music, with all its purported positive effects and benefits on cognitive performance (e.g., Nantais & Schellenberg, 1999; Schellenberg & Hallam, 2005), can reduce the detriments of ISE during a serial word recall task. Accordingly, this study has been designed to contain five auditory conditions: (1) *Instrumental Music-Only*, (2) *Combined* (music with background speech), (3) *Background Speech-Only*, (4) *Scrambled Music with background speech*, and (5) *White Noise*.

Two specific hypotheses follow. First, under the *Background Speech-Only* condition, participants will have the worst recall performance compared to all other conditions, while *Instrumental Music-Only* and *White Noise* conditions will produce the best performances. This hypothesis, if supported, would mean that ISE effects are replicated in a local context, which further qualifies that instrumental music and white noise have less detriments on serial word recall than irrelevant speech does. Second, the *Combined* condition is predicted to yield superior performance in the recall task compared to the *Background Speech-Only* and *Scrambled* conditions would.

Hypothesis 2 addresses the possibility that the instrumental music – with its musical harmony and internal musical structure – may result in a more stable auditory scene for selective processing than the changing-state features of background speech. The *Scrambled* condition consists of the same piece of instrumental music, only rearranged to disrupt its internal musical structure. Hence, the *Combined* condition is predicted to enhance task performance compared to the *Scrambled* condition and *Background Speech-Only* conditions.

Method

Participants

Thirty-five undergraduates from psychology classes in National University of Singapore took part in the study and were awarded course credits for their participation. All participants reported normal hearing.

Design

A 5 [Music Conditions: (1) Music-Only, (2) Background Speech-Only, (3) music with background speech (Combined) versus (4) Scrambled music with background speech, and (5) White Noise] × 2 [Word frequency: high versus low] within-subjects design was used.

Serial Word Recall List

Eighty 4-letter English words were chosen for the 10 word lists (Lim & Yap, 2010); orthographic neighborhood density (held constant at 3.33) and word frequency (high versus low; see Table 1) effects *per se* were *not* expected to emerge in this study (i.e., music condition effects, if any, ought to persist across high and low frequency words).

Table 1: Means and Standard Deviations of Log-frequency for Low and High Frequency Words.

	Log-frequency	
Conditions	*M*	*SD*
Low-frequency	6.61	0.544
High-frequency	11.8	1.230

Stimuli

A total of five auditory conditions were created: *Music-Only*, *Background Speech-Only*, music with background speech (*Combined* condition), *Scrambled music with background speech*, and *White Noise*. The background-speech auditory track was superimposed at the same volume on Bach's Italian Concerto (First Movement) and Haydn's Piano Sonata in E-Flat Major, No. 52. The superimposed tracks were split into sets of 42 seconds each, in order to match the duration of each word list's presentation. This procedure is illustrated in Figure 1.

The same musical track was randomly split and rearranged to create the *Scrambled* condition, therefore maintaining the exact same number of musical notes while disrupting the musical structure. This prevented differences in number of musical notes from producing any differential (unintended) effects in recall performance

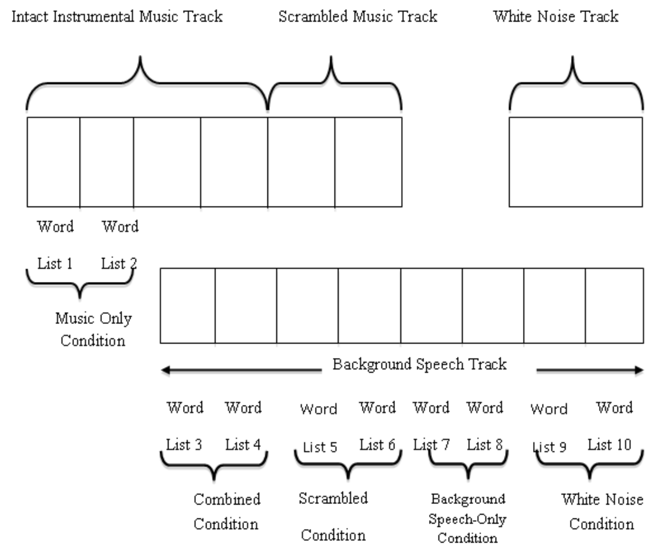

Figure 1: Arrangement of auditory tracks and assignment of word lists.

Manipulation Check

The manipulation check, which asked participants to recall the segments of conversations, was instituted to rule out the possibility that the complete or scrambled music (in the *Combined* and *Scrambled* conditions) was (merely) masking off the background speech track.

Procedure

Participants were presented with the word lists paired with each of the five auditory conditions and were instructed to ignore the background auditory stimuli. Once the list stops, the auditory stimuli paused and the participants were given one minute to recall the words presented. Immediate recall was required after every list using a booklet provided. It was emphasized that only words recalled in the correct position would be scored as correct responses. Exposing participants to different segments specifically controlled for habituation and familiarity effects. In addition, the order of the five different auditory conditions was counterbalanced across the two test sessions.

Results

Analysis of Manipulation Check

The manipulation check was instituted in order to critically rule out the possibility that the music or scrambled music could merely be masking the speech information. Importantly, approximately 73% of the participants recalled more than 2 categories of contents in the background speech. This high recall performance of speech contents constituted important evidence in suggesting that the music tracks did *not* (merely) mask the background speech.

Analysis of Word Recall Performance

A 5✕2 repeated-measures ANOVA was conducted on the percentage of words recalled correctly. The two-way interaction was not significant as earlier predicted, $F(4, 136) = .653$, $MSE = .31$, $p = .626$, and data were subsequently collapsed across word frequency. The main effect of word frequency did not reach significance as well, $F(1, 34) = 1.57$, $MSE = .031$, $p = .219$.

A significant main effect for music conditions emerged, $F(4, 136) = 5.25$, $MSE = .71$, $p = .001$. The irrelevant speech effect was successfully replicated: Post hoc comparisons revealed that percentage of correct recall in the *Background Speech-Only* condition ($M = .554$, $SD = .280$) was significantly lower than recall in *Music-Only* ($M = .713$, $SD = .274$), *Combined* ($M = .664$, $SD = .241$), *Scrambled* ($M = .741$, $SD = .259$) and *White Noise* ($M = .698$, $SD = .275$) conditions. This means that the *Background Speech-Only* condition yielded the worst recall performance compared to all other sound types.

While instrumental music appears to be influential in attenuating the ISE, an intriguing finding was that instrumental music *per se* – specifically its *musical* structure (or *music-ness*) – did not appear to attenuate ISE, due to the fact that the *Scrambled* condition produced significantly higher recall performance than both the *Combined* condition *and Background Speech-Only* condition did. The critical interpretation is that instrumental music attenuated ISE, but musical structure *per se* is not (at least not entirely) responsible for this effect. Figure 2 presents recall performance across conditions.

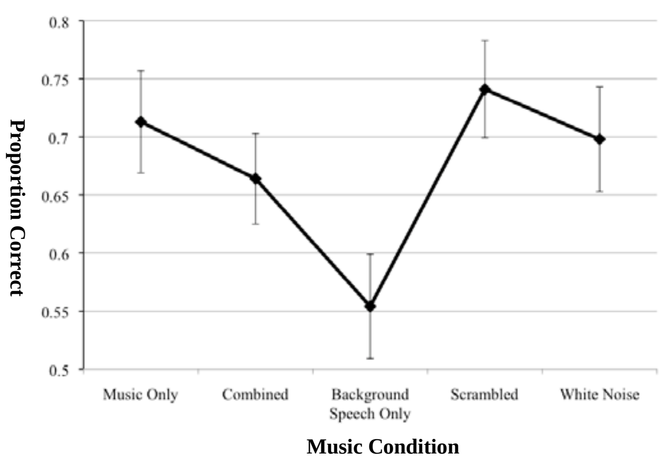

Figure 2: Plots and error bars of mean percentage recall across music conditions. *Background Speech Only* produced the lowest recall, while *Combined* and *Scrambled* yielded significantly higher recall than *Background Speech-Only*.

Discussion

The present results show that the ISE was replicated in a local context. Recall performance in *Background Speech-Only* was the worst compared to that in all other auditory conditions. However, the intriguing finding was the apparent lack of evidence to support the hypothesis that musical structure *per se* can attenuate the ISE (in the *Combined* condition), given the observation that the *Scrambled* condition actually produced better recall scores than both the *Combined* and *Background Speech-Only* conditions did. Auditory masking, albeit a convenient explanation, clearly cannot account for the present data, because of the high percentage of speech contents recalled (73%). Clearly, participants did process the background speech.

Towards a Hybrid Model of Changing States and Attention

Here, we propose a novel hybrid model that combines the attention component from the feature model by Neath (2000) with the changing-state accounts from the O-OER model (Jones, Madden & Miles, 1992) to account for the present data (i.e., improvements in recall performance under both the *Combined* and *Scrambled* conditions compared *to Background Speech-Only* condition).

According to the O-OER model, these changing state features give rise to multiple objects, which interfere with serial processing of the word list compared to a repeated, steady auditory stream. In this study, music did not impose additional processing because it may have less changing features than the irrelevant speech. Therefore, the music tracks are preferred over the irrelevant speech whereby in the *Combined* and *Scrambled* conditions, attention was diverted away from the damaging irrelevant speech. Additional cognitive resources can then be allocated towards the serial word recall task. Ahveninen *et al.* (2011), using multimodal techniques (PET, fMRI, MEG and EEG), found that auditory cortices can selectively deploy attention to segregate relevant sounds from noise, thereby mitigating the detrimental influence of irrelevant speech. In this study, the music track, with less changing-features, makes processing easier, delegating more cognitive resources for the serial word recall task, thereby explaining superior performances in *Combined* and *Scrambled* conditions.

An alternative explanation is that the cumulative presence of the additional auditory stimuli and irrelevant speech in this study led to an increase in distraction levels, resulting in a compensatory increase in attention to the serial recall task. Weissman, Warner and Woldorff (2004) found in their experiment that as the irrelevant stimulus increases in their distraction levels, a compensatory increase in selective attention follows. Therefore, an overall increase in distractibility of auditory stimuli can lead to a compensatory increase in attention, thereby explaining why performances are better in the present *Combined* or *Scrambled* condition.

Summarizing, these findings represent active processing by participants where changing acoustical features of the irrelevant speech and music tracks were compared and the latter (steadier) stream is preferred. Attention is either selectively deployed to the less distracting stream or increased via compensatory mechanisms, allocating more attentional resources towards the serial recall task. Task performance is consequently enhanced.

However, it must be noted that the hybrid model makes the implicit assumption that music has less changing-features than irrelevant speech does, and this model would not particularly aim to differentiate between intact and scrambled music. Therefore, there is a possibility that the present results, where scrambled music yielded better recall performance than did the *Combined* and *Background Speech-Only* conditions, are not (yet) thoroughly accounted for by this model. We next briefly describe (for future work purposes) another property of music that might explain the attenuation of ISE.

Arousal-mood Hypothesis

One particular property of music – arousal – may be promising to explain why scrambled music produced such superior recall performance. The arousal-mood hypothesis by Thompson, Schellenberg, and Husain (2001) argues that the tempo of music is related to arousal while its mode is linked to mood. Music in a major mode corresponds to a happy mood whereas minor mode to a sad mood (Husain, Thompson & Schellenberg, 2001). The re-arrangement of the original Bach and Haydn sonatas music could, in fact, augment the *perceived* tempo in the scrambled track given its now more "staccato-like" (and therefore "rapid") quality (compared to its original unscrambled (and "unrushed") counterparts). The *perceived* faster tempo in the *Scrambled* condition could possibly have produced higher arousal states than did the perceived tempo in the *Combined* condition, which might directly predict recall performance. The view, in a sense, is that the *Scrambled* music then offered listeners with greater cognitive resources (due to heightened states of arousal) to engage in their recall task, than did unscrambled music.

Future Directions

The intriguing finding was that the *Scrambled* condition in fact enhanced recall more than the *Combined* condition did. Since the *musical* structure (i.e., *music-ness*) of the present auditory stimuli did not appear to be (solely) responsible for this attenuation, future research, as recommended above, could explore effects of alternative (e.g., arousal) properties to understand the workings beneath ISE more directly.

Conclusion

This study reports novel data that suggest that ISE can be attenuated (even in a local context) but how that the reason for this attenuation is not (solely) musical structure *per se*. Changing acoustical properties and arousal capabilities of the auditory stimuli may unveil how we might exactly attenuate the ISE. For scrambled music, its arousing properties may potentially attenuate ISE. Therefore, beyond changing-states and selective attention, music's arousing capabilities should be directly investigated in a future study. It is likely that both changing acoustical features and arousal capacities found in music may collectively help attenuate the ISE. These are exciting predictions which would have brought us closer to answering the long-standing question of just why "music" is so capable of offering inoculation against a harsh auditory environment that comprises a host of distractions (e.g., why thousands of students around the world continue to *listen to music* whenever they study).

References

Ahveninen, J., Hämäläinena, M., Jääskeläinenc, I.P., Ahlforsa, S.P., Huang, S., Lina, F.H., Raija, T., Sams, M., Vasiosa, C.P., & Belliveau, J.P. (2011). Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *PNAS*, *108*, 4182–4187.

Baddeley, A. D. (2003b). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.

Colle, H.A., & Welsh, A. (1976) . Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*, 17–32.

Ellermeier, W., & Hellbruck, J. (1998). Is level irrelevant in "irrelevant speech"? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception & Performance*, *24*, 1406–1414.

Farley, L.A., Neath, I., Allbritton, D.W. & Surprenant, A.M. (2007). Irrelevant speech effects and sequential learning. *Memory & Cognition*, *35*, 156–165.

Husain, G., Thompson, W.F., & Schellenberg, E.G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, *20*, 151–171.

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology*, *44A*, 645–669.

Lesiuk, T. (2010). The effect of preferred music on mood and performance in a high-cognitive demand occupation. *Journal of Music Therapy*, *47*, 137–154.

Lim, S. W. H., & Yap, M. J. (2010). *Distributional analyses in visual lexical decision: Orthographic neighborhood density and word frequency effects*. Poster presentation at the 32nd Annual Conference of the Cognitive Science Society, Portland, Oregon, USA; paper published in S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 700–705). Austin, TX: Cognitive Science Society.

Miles, C., Jones, D. M., & Madden, C. A. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 578–584.

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. *Psychological Science*, *10*, 370–373.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, *7*, 403–423.

Schellenberg, E. G., & Hallam, S. (2005). Music listening and cognitive abilities in 10- and 11-year-olds: The Blur effect. *Annals of the New York Academy of Sciences*, *1060*, 202–209.

Tremblay, S., & Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 659–671.

Weissman, D.H., Warner, L.M., & Woldorff, M.G. (2004). The neural mechanisms for minimizing cross-modal distraction. *Journal of Neuroscience*, *24*, 10941–10949.