INTER-NOISE 2018
Impact of Noise Control Engineering
26-29 AUGUST
CHICAGO, ILLINOIS

The effect of "Twinkle Twinkle Little Star" on short-term memory

Munhum Park[a)]
Pavarit Chuprasert
Achcharaphan Kloemwilai
Napat Fahkrajang
Pruch Sawetratanastien
Department of Music Engineering & Multimedia, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, 1 Soi Chalongkrung 1, Ladkrabang, Bangkok 10520, Thailand

ABSTRACT

Irrelevant speech effect (ISE) refers to the decrease in cognitive performance under the influence of a sound which is not related to the task. Typically, ISE has been quantified for speech noise by the rate of errors in recalling visually presented items (serial-recall task). Despite the limited effect size, non-speech sounds have also been shown to disrupt the short-term memory in serial-recall task, of which the spectro-temporal characteristics may easily be adjusted to investigate some aspects of ISE. In the current study, a well-known tune 'Twinkle Twinkle Little Star' was used to create a set of stimuli that differed in three factors: 1) Instrument type (played either on a MIDI piano or in pure tones); 2) number of notes per second (two or four); 3) playing order (original, reverse or random). The results showed some trends that the piano sound disrupts the task slightly more than the pure tone, and so does the four-note version than the two-note. When played in a random order, the error rate was higher than the original-/reverse-order versions, whereas the latter two conditions were almost equal. These trends may be explained in relation to the temporal and spectral variations between sound tokens.

1 INTRODUCTION

Under the influence of task-irrelevant sounds, the performance in various cognitive tasks may decrease, which is often referred to as irrelevant speech or sound effect (ISE)[1]. Where the ISE varies depending on the type of task and the characteristics of sound, the ISE on short-term memory has mostly been studied in experiments in which the participants attempt to recall numbers or items visually presented in a random order in the presence of a speech sound. In these studies, many aspects of the effect were correlated to the nature of stimulus, testing several hypotheses, of which the changing-state hypothesis appears to be most robust[2].

[a)] email: munhum.pa@kmitl.ac.th

The ISE incurred by non-speech sounds, however, has not drawn much attention so far, partly because the size of effect is much smaller than speech sounds. In earlier studies, pure tones were found to disrupt the serial recall task as much as spoken syllables[3]. Salamé and Baddeley[4] carried out a series of experiments using vocal/instrumental music together with silence and speech, where the vocal music was consistently found more disruptive than the instrumental music, of which the latter was not significantly different from silence. Where the meaning of speech was found not to influence the ISE[1], Nittono[5] showed that the recording of an instrumental music played forward disrupted the task more than that played backward, suggesting that a higher-level cognitive process may be involved, taking into account the meaning of music.

In the current study, the non-speech ISE was investigated by making several variations of a well-known tune, 'Twinkle Twinkle Little Star.' The tune was played by using a MIDI piano or by pure tones of which the envelope was identical to that of the piano sound. Also, the tune was rearranged so that the number of notes per measure differed by half while the modified tune was still recognized as the original tune. Lastly, the tune was played either in the original, reverse or random order in order to investigate the influence of the musical meaning on the ISE. Given these 12 variations of the tune, the ISE was quantified with reference to silence and a speech sound.

Section 2 describes the stimuli and the test design in details, followed by the results and discussions in sections 3 and 4, and a summary is given in section 5.

2 METHODS

2.1 Task

A typical serial recall task was used in the current study to quantify the effect of sounds on short-term memory. One trial began with nine numbers from 1 to 9 appearing on the computer screen every 1 second in a random order. After a 10-second retention period, the participant must use an on-screen 3-by-3 keypad to recall the nine numbers in the correct order. During each trial, one of the 14 sound conditions (see section 2.2) was presented by using a sound card (Steinberg UR22) and headphones (Shure SRH440), which the participant wore throughout the experiment (except for break times). The number of misrecalled digits was counted as the error rate for each trial. Participants were given a small amount of honorarium. For a more detailed description of the task and the test interface, readers are referred to Park et al.[6] The experiment was carried out in a recording studio, which was isolated well from external noise.

2.2 Stimulus

A set of 13 sound stimuli were prepared, which included 12 variations of "Twinkle Twinkle Little Star" and a speech sound (spoken in Thai). The Thai speech sequence was a female monologue recorded from a radio news channel. The 12 variations of Twinkle Twinkle Little Star (not to be confused with the popular arrangement by Mozart) differed in three factors: Instrument (piano or tone), the number of notes per measure (two or four; abbreviated as NPM, hereinafter), and the playing order (original, reverse and random) as listed in Table 1.

Table 1. Twelve variations of Twinkle Twinkle Little Star

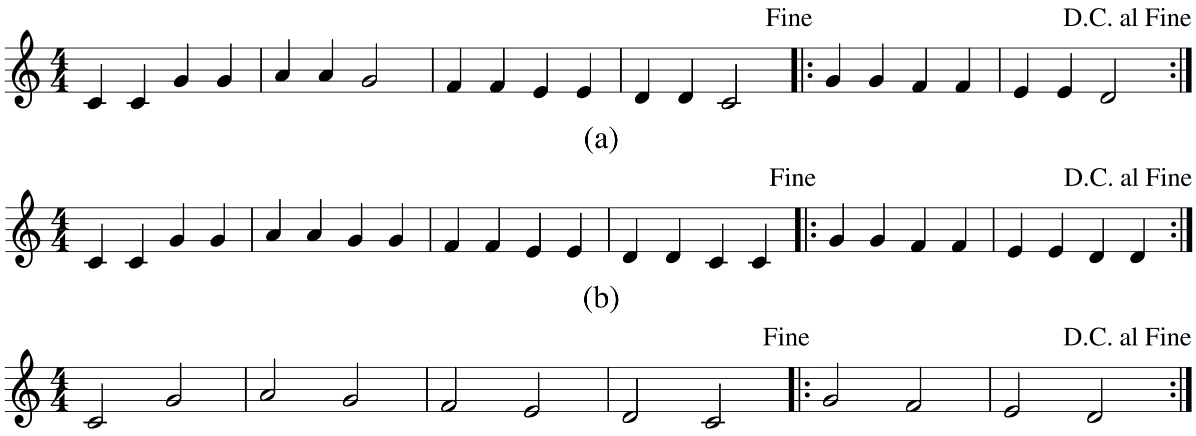
Acronym for stimulus	Instrument	Number of notes per measure (NPM)	Playing order
piano_four_orig	piano	4	original
piano_four_rand	piano	4	random
piano_four_rev	piano	4	reverse
piano_two_orig	piano	2	original
piano_two_rand	piano	2	random
piano_two_rev	piano	2	reverse
tone_four_orig	tone	4	original
tone_four_rand	tone	4	random
tone_four_rev	tone	4	reverse
tone_two_orig	tone	2	original
tone_two_rand	tone	2	random
tone_two_rev	tone	2	reverse

- Instrument (piano or tone): For piano, a MIDI software (Logic Pro X) was used to record either 0.25-s or 0.5-s long notes from C to A required to play the tune, where the same velocity was maintained for all notes. For tone, an average envelope of waveform was first extracted from the piano sounds, within which a pure tone sound only of the fundamental frequency from C to A was filled. By comparing the two instrument types, the influence of the presence of higher harmonics on ISE may be studied.

- Number of notes per measure or NPM (2 or 4): In the melody of Twinkle Twinkle Little Star, the same quarter note is always played twice except for the half notes appearing every two measures (see Fig. 1a). For the four-note version, the tune was slightly modified by splitting all half notes into two quarter notes, so that all measures consist only of quarter notes (see Fig. 1b). Also, the two-note version was created by combining pairs of quarter notes in the original tune into half notes as shown in Fig. 1c. Although modified, both four-note and two-note versions can easily be recognized as the well-known tune when played in the original order. For both versions, the tune was played at the speed of one measure per second. In other words, the length of each note was 0.25 second for the four-note version, and 0.5 second for the two-note version. By comparing these two versions of the tune, the influence of the temporal structure on ISE may be studied.

- Playing order (original, reverse and random): The two versions of the tune differing in NPM were played either in the original order or in the reversed order. In addition, the notes in each version were randomly ordered and played without the same notes played successively. Obviously, the reverse-order version may not be recognized as Twinkle Twinkle Little Star, which would however sound much more “musical” than the random-order. By comparing the three different orders, the influence of the musical meaning on ISE may be studied.

In addition to the 13 ‘sound’ conditions, the participants also performed the task in silence (but wearing the headphones for the consistency). Therefore, a total of 14 conditions (13 sound and 1 silence) were used in the current study, where the error rates under the silence and the speech condition may be the lower and the upper boundaries, respectively.



(c)

Figure 1. (a) The original tune of Twinkle Twinkle Little Star. (b) The four-note version. (c) The two-note version.

2.3 Test design

One test session consisted of 4 blocks of 21 trials, which required, including 2-minute breaks between blocks, about an hour to complete. Some training trials were provided so that the participants may get used to the task. Table 2 shows the number of training and test trials in each block. Given 14 test conditions, each condition could be repeated 5 times in a session, and therefore, 70 test trials were randomly presented in each session. All participants were asked to complete 4 sessions (one session per day) not necessarily on consecutive days but within two weeks. At the end of the experiment, 20 error rates were available per condition and per participant for the analysis.

Table 2. The number of trials for training and actual test in each block.

	Block 1	Block 2	Block 3	Block 4
Training trials	5	1	1	1
Test trials	16	20	20	14

2.4 Analysis

Data were analyzed by using repeated-measure analysis of variance (ANOVA) and multiple linear regression. For the ANOVA, error rates were first averaged for all 20 trials per condition per participant, and this modified data set was analyzed for the effect of condition (with 14 levels). For the post-hoc analysis, paired t-tests were carried out with and without Bonferroni correction.

In order to further investigate the effect of the type of instrument, NPM and the playing order, the data for the silence and the speech conditions were first removed, and the raw error rates (without averaging as in the ANOVA) under the remaining 12 conditions were analyzed by using a multiple linear regression model. Together with 'Instrument,' 'NPM,' and 'order,' the unique number given to each participant (participant ID) and the session number (1~4; session ID) were also regarded as independent variables in the model. All statistical analyses were carried out at the significance level of 0.05 either on Python or R[7].

3. RESULTS

15 undergraduate students [7 female (mean age: 19.4) and 8 male (mean age: 19.1)] took part in the experiment, all of who speak Thai as their mother tongue. No hearing problem was reported during the pre-test interview.

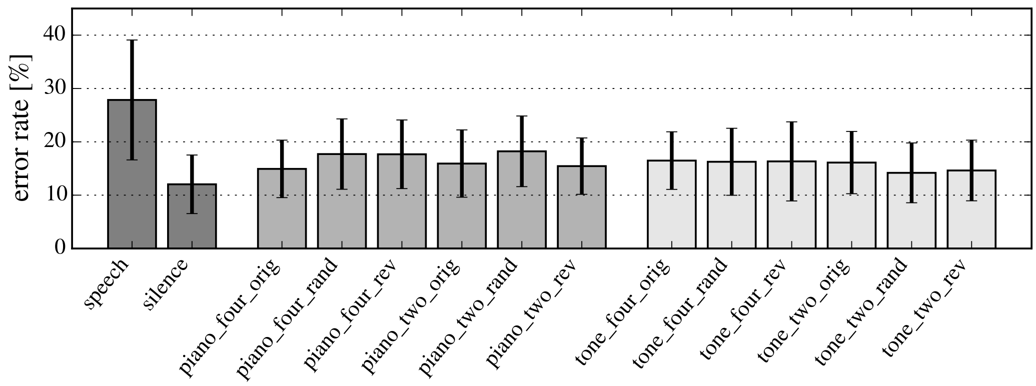


Figure 2. Overview of the error rates with means and confidence intervals shown for all 14 conditions.

Figure 2 shows an overview of the test results with the means and confidence intervals of the averaged error rates for all 14 conditions. The effect of the sound condition was found to be significant by the repeated-measure ANOVA [$F(3.65, 51.1) = 5.88$, $p<0.001$; degrees of freedom corrected for the sphericity violation]. When paired t-tests were carried out for all pairs of the sound conditions (without Bonferroni correction), the error rate under speech was higher than the other 13 conditions ($p\leq0.026$). The silence condition was shown to produce the error rate lower than most conditions ($p\leq0.047$), whereas the comparison was not significant with piano_two_rev, tone_two_rand, tone_two_rev, and tone_four_rev. As for those pairs consisting of the conditions other than silence and speech, five comparisons were found to be significant as listed in Table 3. When the Bonferroni correction was used, however, none of the t-tests reached the significance level, mainly because of the large number of pairwise comparisons.

Table 3. The pairs of conditions (except for silence and speech) of which the difference in the error rate was significant.

piano_four_orig vs. piano_two_rand	piano_four_orig vs. piano_four_rev	piano_four_rev vs. tone_two_rev	piano_four_rev vs. tone_two_rand	piano_two_rand vs. tone_two_rand
p=0.045	*p=0.021*	*p=0.009*	*p=0.008*	*p=0.048*

The results of the multiple linear regression showed that the participant ID and the session ID were significant factors that affected the error rates as shown in Table 3. However, the instrument type, NPM and the playing order did not reach the significance level. The adjusted R^2 value associated with the current linear model was only 0.21.

Table 4. Independent variables of the linear model fitted to the error rates. Adjusted $R^2=0.21$.

Session ID	Participant ID	Instrument	NPM	Order
*P<0.001**	*P<0.001**	*p=0.170*	*p=0.258*	*p=0.679*

As presented above, the three main factors of the experiment did not significantly affect the error rates. Nevertheless, some trends could be observed as shown in Fig. 3, where the corrected error rate (error rates predicted by the linear model when the other factors are fixed to the most common values) was higher for the piano than for the tone, and for the four-note than for the two-note. Randomly shuffling the original order seems to have slightly increased the error rate compared to the original or the reverse order (see Fig. 3c).

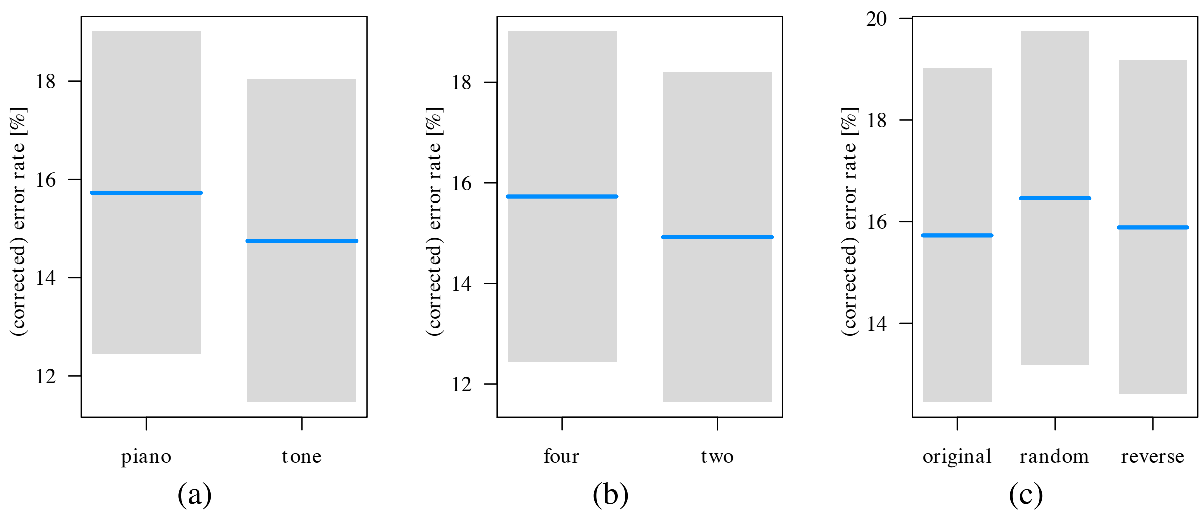


Figure 3. With the other two factors held constant, the linear model in the current study compared the (corrected) error rates for the three factors: (a) Instrument type, (b) number of notes per measure, and (c) playing order. The lines and the boxes in the figures indicate the means and the 95-% confidence intervals.

4. DISCUSSIONS

As expected from the test design phase, the error rates associated with speech and silence were shown to be the upper and lower boundaries (see Fig. 2). The 12 variations of Twinkle Twinkle Little Star disrupted the serial recall task significantly less than speech, supported by the results of the paired t-tests. On the other hand, the same statistical tests showed that the difference is not significant at least between the silence condition and 4 variations (piano_two_rev, tone_two_rand, tone_two_rev, and tone_four_rev). It appears that the musical sounds used in the current study disrupted the serial recall task only slightly more than the silence condition in clear contrast to the effect of speech sound, which agrees well with the observations of some previous studies[3,4].

When compared only between the 12 variations (without silence and speech) using paired t-tests, no consistent influence of either the instrument type, NPM or the playing order could be observed as shown in Table 3, although the difference between the listed pairs was significant (without the Bonferroni correction). The results of the multiple linear regression were not significant for these three factors either, but some interesting trends could be observed. For example, the piano increased the ISE slightly more than the tone (see Fig. 3a), and so did the four-note version compared to the two-note (Fig. 3b). The latter may be related to the increased rate of modulation in the four-note version, thus resulting in a higher error rate compared to the two-note version. The former difference between the piano and the pure tone may be related to the difference in both spectral and temporal changes, as the temporal changes take place in a wider frequency band in the piano sound than in the tone.

The effect of the playing order was least significant in the regression analysis compared to the other independent variables, as shown in Table 4. More details may be observed in Fig. 3c, where the error rates were similar between the original and the reversed versions, implying that the participants' recognition of the tune did not increase or decrease the performance in serial-recall task. In other words, the 'meaning' of the music hardly influenced the ISE, as is the case with the meaning of speech sounds. This result is somewhat contradictory to the findings by Nittono[5] where the forward playback of a music recording disrupted the task more than the backward playback. Since playing back a recording backward changes the temporal structure of sounds, for example, by making the release come before the attack of each sound, the 'perceived' modulation depth may be less than that of the forward playback, which may have decreased the ISE.

The same figure (Fig. 3c) however shows that the random version slightly increased the error rate (although not significant) compared to the original and the reversed, which may be explained by the change in modulation: As mentioned in section 2.2, the random version was created by shuffling all notes so that no successive notes play at the same pitch, in which case the degree of change in the sound would increase, especially in the four-note version. (The two-note version is not affected by shuffling, where all neighboring notes played different tones already in the original order.) So, the increase in modulation may have increased the error rate under the random-order condition compared to the original or reversed orders.

Despite some noteworthy observations, differences between the levels of the three factors (instrument, NPM and order) were not statistically significant, mainly because the effect size of the musical sounds was so small compared to the speech sound. Given the small effect size, it was and would be very challenging to detect the individual effect of different factors, even if the sample size were larger. For example, the difference in the corrected error rate was only about 1 percent between different levels of the three factors (see Fig. 3), which would hardly have any meaningful implication in the real world. It is likely that stimuli consisting of multiple streams of sounds (e.g., played by different instruments possibly asynchronously) may incur the ISE to a similar level of speech, in which case the effects of the playing order and other factors considered in the current study might be easier to detect.

On a side note, the participants improved their short-term memory skills day by day, as clearly indicated by the significant effect of session ID (see Table 4 and Fig. 4; also see Park et al.[6]). Within the test design used in the current study, the error rates under each condition were obtained in four sessions, and therefore, the effect of session ID must have unnecessarily increased the variance in the data. Follow-up experiments may be designed so that only a small number of conditions are compared within a single session.

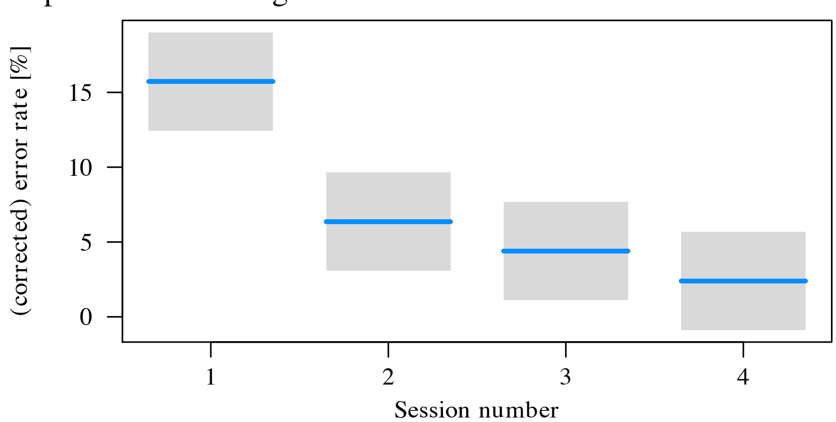


Figure 4. Error rate decreases as the participants improve their skills to perform the serial recall task.

5. SUMMARY

In the current study, the irrelevant speech effect was studied for non-speech sounds by using a well-known tune, 'Twinkle Twinkle Little Star,' of which 12 variations were created by differentiating three factors: The instrument type, the number of notes per measure, and the playing order. In the serial-recall task, participants made slightly more errors when the tune was played on piano than on a pure-tone instrument, and when more notes were played per second. Compared between different playing orders, the random-order version resulted in a slightly higher error rate than the original and the reversed. Given the results of the experiment, it appears that the temporal and spectral variations in sound tokens may influence the ISE of non-speech sounds, whereas the meaning (or the recognition) of the sound may not, which is consistent with the ISE of speech sounds. Nevertheless, the effect size was limited, and therefore, most comparisons made between non-speech conditions were not statistically significant.

6 ACKNOWLEDGEMENTS

Authors thank the participants of the experiment at the Department of Music Engineering and Multimedia, KMITL.

7 REFERENCES

1. Banbury, S. P., Macken, W. J., Tremblay, S. & Jones, D. M. Auditory distraction and short-term memory: Phenomena and practical implications. *Hum. Factors* **43,** 12–29 (2001).
2. Ellermeier, W. & Hellbruck, J. Is level irrelevant in 'Irrelevant speech'? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *J. Exp. Psychol.-Hum. Percept. Perform.* **24,** 1406–1414 (1998).
3. Jones, D. M. & Macken, W. J. Irrelevant Tones Produce an Irrelevant Speech Effect - Implications for Phonological Coding in Working Memory. *J. Exp. Psychol.-Learn. Mem. Cogn.* **19,** 369–381 (1993).
4. Salame, P. & Baddeleyb, A. Effects of background music on phonological short-term memory. *Q. J. Exp. Psychol. Sect. A* **41,** 107–122 (1989).
5. Nittono, H. Background instrumental music and serial recall. *Percept. Mot. Skills* **84,** 1307–1313 (1997).
6. Park, M., Kohlrausch, A. & van Leest, A. Irrelevant speech effect under stationary and adaptive masking conditions. *J. Acoust. Soc. Am.* **134,** 1970–1981 (2013).
7. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2018).