Applied Acoustics 124 (2017) 30–37

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Effect of word familiarity on word intelligibility of four continuous words under long-path echo conditions

Zhenglie Cui [a], Shuichi Sakamoto [a,*], Masayuki Morimoto [a], Yôiti Suzuki [a], Hayato Sato [b]

[a] Research Institute of Electrical Communications, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi 980-8577, Japan
[b] Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan

ARTICLE INFO

Article history:
Received 14 October 2016
Received in revised form 15 January 2017
Accepted 2 February 2017

Keywords:
Word familiarity
Word intelligibility
Four continuous words
Long-path echo
Open-air public address systems

ABSTRACT

The massive earthquake that occurred on 11 March 2011 in Japan demonstrated that the intelligibility of speech presented over mass notification sound systems is often significantly degraded by long-path echoes. This study examines the effect of word familiarity on speech intelligibility in the presence of long-path echoes, in order to increase speech intelligibility in such systems. We performed two experiments using sets of four sequentially connected words (quadruplets), in place of an actual sentence. In Experiment 1, we investigated word intelligibility in the presence of simulated long-path echoes for quadruplets consisting of words with the same word familiarity rank. The results indicated that the intelligibility of high-familiarity words is higher than that of low-familiarity words, irrespective of the number of simulated long-path echoes. In Experiment 2, quadruplets with mixed word familiarity were used to investigate intelligibility under more realistic conditions. The results of Experiment 2 demonstrate that the intelligibility of high-familiarity words is higher than that of low-familiarity words under long-path echo conditions, even when high- and low-familiarity words coexist in one quadruplet. These facts show that high-familiarity words are more robust against the influence of long-path echoes than low-familiarity words, strongly suggesting that announcements presented from mass notification sound systems should consist of high familiarity words as much as possible.

1. Introduction

The massive earthquake and high tidal wave that struck Japan on 11 March 2011 reinforced the importance of mass notification systems. Among various systems, the governmental mass notification sound system that covers wide coastal areas of Japan has a significant advantage: it can effectively convey emergency announcements over wide areas and does not require message recipients to use any special equipment. This system broadcasts messages from local governments to citizens via radio communications and public address systems installed on towers. This system, however, confronted us with a serious problem: direct messages to citizens in emergency evacuations are often insufficiently intelligible. In fact, a large governmental questionnaire indicated that 20% of the citizens judged the messages to be unintelligible [1].

Long-path echoes are the most significant cause of reduced speech intelligibility in messages conveyed by mass notification sound systems [1]. That is, in open-air public address systems, echoes with delay times that are much longer than those observed in rooms are generated by reflections from hilly terrain, buildings, and other obstructions. Moreover, neighboring public address towers presenting the same message may cause long-path echoes, with delay times corresponding to the differences in the sound propagation paths between the towers and the listening point.

Nonetheless, the manner in which long-path echoes interfere with speech understanding has not been well investigated. Before the earthquake, Toida investigated the effect of long-path echoes on outdoor speech intelligibility [2–4] and proposed a method to predict the speech intelligibility of Japanese mono-syllables, but not the intelligibility of words and sentences that are suitable for practical use [4]. After the earthquake, Sato et al. measured word intelligibility in several open-air environments and in various simulated long-path echo conditions in a laboratory, and found that the speech transmission index (STI) has potential as an objective measure to estimate speech intelligibility, even in the presence of long-path echoes [5]. However, to the best of our knowledge, no studies have proposed methods to increase the robustness of open-air public address systems in terms of speech intelligibility in the presence of long-path echoes. To cope with this problem,

* Corresponding author.
E-mail address: saka@ais.riec.tohoku.ac.jp (S. Sakamoto).

http://dx.doi.org/10.1016/j.apacoust.2017.02.001

the authors focus on the familiarity of words used to compose messages presented from open-air public address systems.

This study considers the familiarity of words that compose announcements presented from loudspeakers as a measure to decrease interference by long-path echoes. "Word familiarity" is a subjective variable that indicates how familiar native speakers are with the word. We used the familiarity values measured by Amano et al. [6] for about eighty thousand Japanese words, i.e., all the entry words in a popular Japanese dictionary [6]. Familiarity was rated from 1.0 (minimum) to 7.0 (maximum). Sakamoto et al. [7] showed that word intelligibility increases as the familiarity increases in both quiet and noisy environments. Sato et al. [8] demonstrated that word intelligibility increases as word familiarity increases in noisy and reverberant environments. Morimoto et al. [9] indicated that listening difficulty decreases and word intelligibility increases when word familiarity increases in reverberant environments. These results clearly indicate that high-familiarity words are robustly intelligible under various severe listening conditions. We can expect, therefore, that high-familiarity words presented with long-path echoes will be more robust, in terms of intelligibility, than low-familiarity words.

The purpose of this study is to clarify the effect of word familiarity on speech intelligibility in the presence of long-path echoes. In this study, we performed two experiments using sets of four sequentially connected words, in place of an actual sentence. Hereafter, a series of four continuous words is called a quadruplet. In the first and second experiments, respectively, we used quadruplets containing words with equal familiarity ranks and quadruplets containing words with mixed fmiliarity ranks.

2. Experiment 1: Intelligibility of four - continuous words with the same word familiarity

2.1. Apparatus

The experiment was conducted in a soundproof room at the Research Institute of Electrical Communications, Tohoku University. Acoustic stimuli were presented diotically to both ears using headphones (Sennheiser HDA-200) through an audio interface (Cakewalk UA-25EX) connected to a notebook computer.

2.2. Test words and test sound fields

The test words were selected from a familiarity-controlled word list called FW07 [10]. The word list consists of four word familiarity ranks as follows: highest (7.0–5.5), second-highest (5.5–4.0), second-lowest (4.0–2.5), and lowest (2.5–1.0). Each rank consists of 20 lists, and each list contains 20 words, i.e., each rank includes 400 words. All the words have four morae.[1] The words were spoken by a trained female speaker, and recorded in a studio. The quadruplets used in this experiment were composed of either (1) words with the highest familiarity ranks (7.0–5.5) or (2) words with the second-highest familiarity ranks (5.5–4.0). All the words in the database are scored over a range of 7.0–1.0 in familiarity, grouped into four ranks: 7.0–5.5, 5.5–4.0, 4.0–2.5, and 2.5–1.0. However, most words presented via actual mass-notification systems are in the familiarity range between 7.0 and 4.0 [11]. Consequently, only the words within the familiarity ranges of 7.0–5.5 and 5.5–4.0 were selected in the present study. Hereafter in this paper, the former is called high-familiarity stimuli and the latter low-familiarity stimuli.

Fig. 1 diagrams the time patterns of presented words for seven types of simulated sound fields with different echo patterns used in the experiment. Fig. 2 depicts the examples of the time waveform and spectrogram in conditions A and C. Here, a preceding sound means the quadruplet presented at first to simulate the speech sound arriving at a listening point via the shortest path, and a following sound means the quadruplet presented with a lag time from the preceding sound to simulate a long path echo.

Sound field A consists of only a preceding sound with no following speech sound (simulated speech sound without any long-path echo), B, C, and D consist of a preceding sound and a single following sound (speech sound with a single simulated long-path echo), and E, F, and G consist of a preceding sound and following sounds (speech sound with two simulated long-path echoes). A cluster of symbols (circles, squares, triangles, or rhomboids) represents an individual word and one symbol represents each mora of each word. The delay time between the preceding sound and the following sound was set to multiples of 375.0 ms. This length was decided on the basis of the average word length of the 4-mora-words used in the experiment. The average and standard deviations of word lengths for all words used in the experiments were 773 ms and 53 ms, respectively, and the calculated word length was rounded to 750 ms. That is, 375.0 ms corresponds to half the average word length, and this value was used as the basic unit for the delay times of the following sounds in all experimental conditions. For example, in condition E, a preceding sound, a following sound with a time delay of 375 ms and another with a time delay of 750 ms were presented to listeners. The preceding sound and following sound(s) ended at the same time, as shown in Fig. 1.

All Japanese words consist of units called "mora." Mora is the basic unit of the Japanese language rather than syllable and is similar to rhythm beat in music as explained previously. Therefore, the lengths of moras distribute in a narrow range if the speech rate is constant. As written in the manuscript, all the words in the database, and thus those used in this study, consist of four moras. As a consequence, the standard deviation of the word lengths is only 53 ms while the mean is 773 ms. Moreover, since the words consist of four moras, the standard deviation of the mora lengths in the words becomes 26.5 ms, corresponding to around one-eighth of the mean of the mora length, resulting in very consistent lengths of the quadruplet used. Therefore, the stimuli can be regarded as having enough similar lengths so that their time pattern can be represented by the time pattern model shown in Fig. 1.

The sound pressure levels of the words were measured by 1/2 inch microphones (B&K 4192) attached to an artificial ear (B&K 4153). The A-weighted sound pressure level was set to 60 dB (L_{Aeq}). In this study, the sound pressure levels of the sounds were identical. For this experiment, we assumed the situation that the same sounds are presented with certain delay from different loudspeakers located at the same distance from the listening point. In this situation, two sounds arrive at the listening point with the same sound pressure level. This can be considered as the worst case; once there is a difference between the sound pressure levels of preceding and following sounds, listeners can easily hear the sound with the larger sound pressure level [12]. Therefore, in this study, the sound pressure level of the following sound was set at the same level as that of the preceding sound.

2.3. Listeners

Six male listeners and one female listener (ranging from 20–24 years old) with normal hearing sensitivity participated in the experiment.

[1] Unit to count syllable-like structures of the Japanese language [7]. For example, the Japanese word "pan" (bread) consists of two morae of "pa" and "n" although this is a mono-syllabic word.

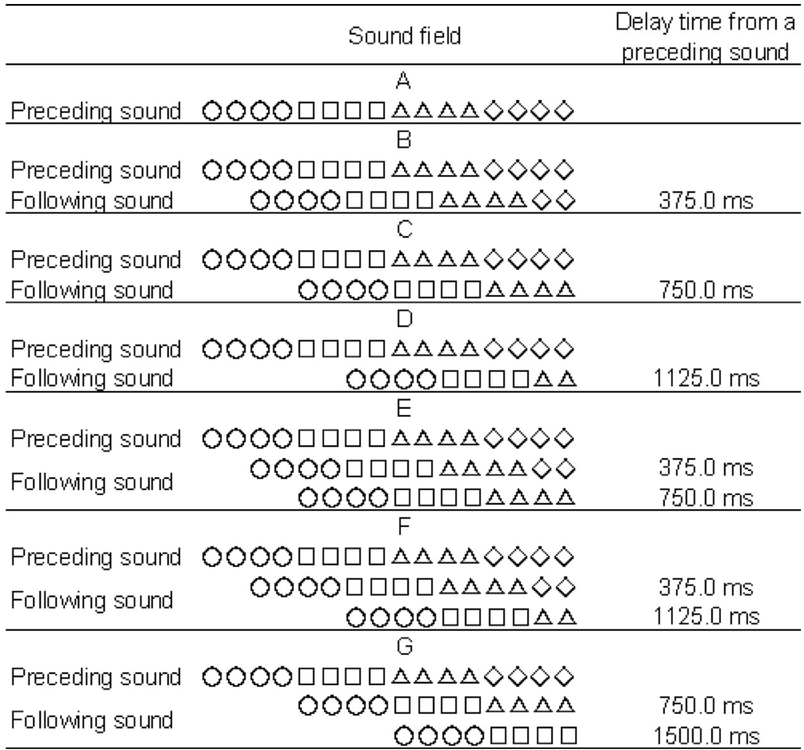
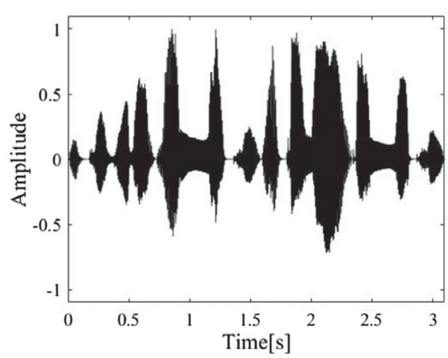
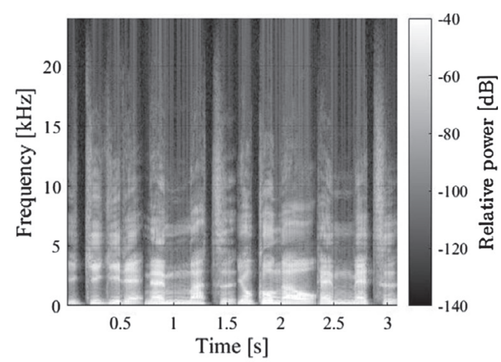


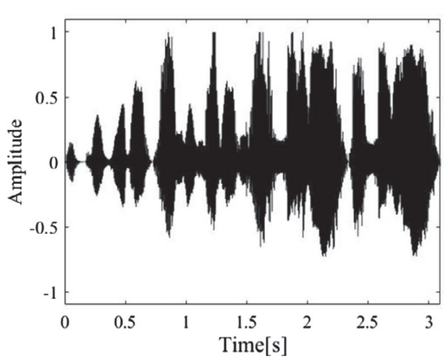
Fig. 1. Pattern diagram of seven sound fields used in Experiment 1. A cluster of the same four symbols (circles, squares, triangles, or rhomboids) represents an individual four-mora word and one symbol represents each mora of each word.



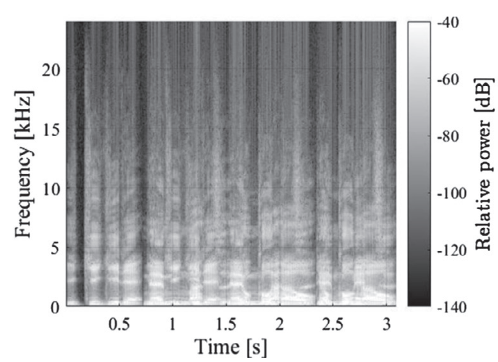
(a) Time waveform (condition A).



(b) Spectrogram (condition A).



(c)Time waveform (condition C).



(d) Spectrogram (condition C).

Fig. 2. Examples of time waveforms and spectrograms of the stimuli in conditions A and C.

2.4. Procedure

Two twenty word lists associated with either the high or the low rank familiarity were used. Twenty words were included in each list as mentioned above.

For each of the high and low word familiarity ranks, 20 word lists were divided into five sets of four lists. To produce a quadruplet, the four lists in one set were, respectively, assigned to the positions (the first to the fourth) of the quadruplet. Then, one word was selected randomly from each of the four word lists and the four words were connected sequentially to produce a quadruplet. The used words were omitted and this process was repeated 20 times to obtain a speech unit consisting of 20 quadruplets. This procedure was applied to the five sets. As a result, five speech units in total, consisting of 20 quadruplets each, were obtained. Because only five sets were available to test seven types of sound fields (A to G), we assigned the seven listeners to each experimental condition as shown in Table 1, which shows the assignment of the speech unit/sound field combinations for listeners, in this manner, a word appears only once for a specific listener. "L" in the table indicates each listener. Each listener participated in only five of the seven types of sound fields and listened to a different speech unit in each sound field. As a result, a listener did not listen to the same word more than once. The number of quadruplets used to test each sound field was 100 (20 quadruplets $\times$ 5 listeners). The experiment with the high and low word familiarity ranks was conducted separately. All listeners participated in both experiments.

The experiment with the high word familiarity was performed first for all listeners. In one session, stimuli with the same sound field were presented. Here, the 20 quadruplets were presented randomly. The inter-stimulus interval for presentation of the quadruplets was set to 15 s. Listeners were asked to write down each quadruplet exactly as they listened to during the 15 s interval. The order of the speech units was not randomized but presented from unit 1–5 for all listeners. The total test duration for each listener was approximately two hours, including short breaks between sessions.

2.5. Results and discussion

In total, 80 test words included in 20 quadruplets were presented for each listener for each sound field. The percentage of test words answered correctly was calculated as a word intelligibility score for each listener, for each sound field. Fig. 3 shows the word intelligibility scores averaged over the five listeners. Error bars denote the standard deviations. For any word familiarity rank, the word intelligibility score decreased by adding a following sound. Moreover, adding another following sound further decreased the word intelligibility score. Nevertheless, for all sounds fields, word intelligibility scores for high-familiarity words were higher than those for low-familiarity words.

Two-way ANOVA was carried out using sound field and word familiarity as factors. The result revealed the significant effects of both sound field ($F(6,28) = 41.08$, $p < .001$) and word familiarity ($F(1,28) = 57.09$, $p < .001$). Multiple comparisons (Ryan's method, $p < .05$) on the main effect of sound fields revealed that the word intelligibility of the condition with a preceding sound only (sound field A) is significantly higher than the word intelligibilities of all of other sound fields (B-G), and that word intelligibility scores for sound fields composed of a preceding sound and a single following sound (sound fields B-D) were significantly higher than those for sound fields composed of a preceding sound and two following sounds (sound fields E-G).

Fig. 4 shows word intelligibility scores for each word position in a quadruplet for each sound field, in order to show the effect of word familiarity on word intelligibility for each word position in a quadruplet. For all sound fields and all word positions, the intelligibility of high-familiarity words is higher than that of low-familiarity words, except for the first position of sound field D and the fourth position of sound field F. Three-way ANOVA was carried out using sound field, word familiarity, and word position as factors. The result reveals significant main effects for all factors: sound field ($F(6,28) = 41.08$, $p < .001$), word familiarity ($F(1,28) = 57.09$, $p < .001$), and word position ($F(3,84) = 348.70$, $p < .001$). This result clearly indicates that the intelligibility of high-familiarity words is higher than that of low-familiarity words.

In summary, the results of Experiment 1 demonstrate that long-path echoes degrade word intelligibility, and that intelligibility decreases with an increase in the number of echoes. Nevertheless, under long-path echo conditions, the intelligibility of high-familiarity words is higher than that of low-familiarity words, regardless of their position in a quadruplet. These facts suggest that high-familiarity words are more robust against the influence of long-pass echoes than those low-familiarity words in open-air environments.

3. Experiment 2: Intelligibility of four continuous words with mixed word familiarity

The results of Experiment 1 suggest that high-familiarity words are robust against the influence of long-pass echoes in comparison with low-familiarity words. However, the familiarity ranks of words used for the preceding sound and for the following sounds were identical in Experiment 1. The overlapping of words with different familiarities may affect the words intelligibility, and these effects should be analyzed. Moreover, in actual situations, it is impossible to compose announcements that use only high-familiarity words. In Experiment 2, the word intelligibility of quadruplets with mixed word familiarity was measured to investigate the influence of long-path echoes under more realistic conditions than in Experiment 1.

3.1. Apparatus, test words, and sound fields

The apparatus was identical to that used in Experiment 1. Quadruplets used in the experiment were composed of high-familiarity and low-familiarity words appearing alternately. The words appeared in either high-low order (HLHL) or low-high order (LHLH). As in Experiment 1, high familiarity refers to the highest familiarity rank (7.0–5.5) and low familiarity refers to the second-highest (5.5–4.0).

Fig. 5 diagrams the time patterns of presented words for simulated sound fields A, C, and G. Because the results of Experiment 1 indicated no significant differences among sound fields B, C, and D (conditions with a single echo), and among sound fields E, F, and G (conditions with two echoes), only three long-path echo conditions (A, C, G) were selected. From the figure, it can be seen that a high-familiarity word overlaps completely with a low-familiarity word and vice versa at the same word position.

In the same manner as in Experiment 1, the sound pressure level of words was measured by 1/2 in. microphones (B&K 4192) attached to an artificial ear (B&K 4153). The A-weighted sound pressure level was set to 60 dB (L_{Aeq}).

3.2. Listeners

Five male listeners and one female listener (ranging from 20–24 years old) with normal hearing sensitivity participated in the experiment. None of them had participated in Experiment 1.

Table 1
Assignment of listeners to the speech unit/sound field combinations in Experiment 1.

Speech unit	Sound field						
	A	B	C	D	E	F	G
I	L1	L2	L3	L4	L5	L6	L7
II	L7	L1	L2	L3	L4	L5	L6
III	L6	L7	L1	L2	L3	L4	L5
IV	L5	L6	L7	L1	L2	L3	L4
V	L4	L5	L6	L7	L1	L2	L3

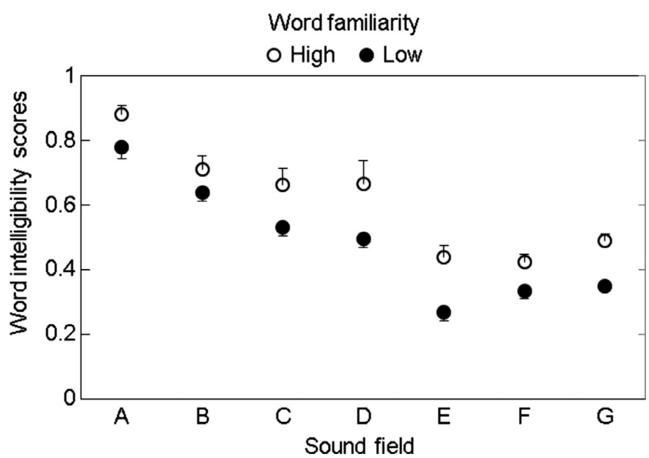


Fig. 3. Word intelligibility scores averaged over seven participants.

3.3. Procedure

Quadruplets used in Experiment 2 were generated using a method similar to that used in Experiment 1. For each of the high and low word familiarity ranks, 20 word lists were divided into ten sets of two lists. To produce a quadruplet with an appearance order of HLHL (high-low), four word lists (i.e., one set of two high-familiarity word lists and one set of two low-familiarity word lists) were assigned to the positions (the first to the fourth) of the quadruplet in high-low-high-low (HLHL) order. Then, one word was selected randomly from each of the four word lists and the four words were connected sequentially. The used words were omitted and this process was repeated 20 times to obtain a speech unit consisting of 20 quadruplets. This procedure was applied to the ten sets. As a result, ten speech units consisting of 20 quadruplets in HLHL order were obtained. Using a similar procedure, ten speech units consisting of 20 quadruplets each in low-high-low-high (LHLH) order were obtained.

Table 2 shows the assignment of participants to the speech unit/sound field combinations. "L" in the table indicates each listener. L1 to L3 participated in the experiment that used HLHL quadruplets and L4 to L6 participated in the experiment that used LHLH quadruplets. As shown in the table, each listener participated in ten conditions to listen to a different speech unit in each condition. As a result, a specific listener did not listen to the same word more than once. For each order (high-low, low-high) the number of quadruplets used to test each sound field was 60 (20 quadruplets $\times$ 3 listeners).

In one session, stimuli with the same sound field were presented. Moreover, the 20 quadruplets presented in one session were presented randomly. The interval between presented quadruplets was 15 s. Listeners were asked to write down each quadruplet exactly as they listened to during the 15 s interval. The total test duration for each participant was approximately two hours, including short breaks between sessions.

3.4. Results and discussion

Fig. 6 shows the word intelligibility scores for each sound field with quadruplets having HLHL and LHLH orders, by pooling the familiarity and word position. The figure shows word intelligibility scores averaged over three listeners. The word intelligibility of sound field A, which consists of only a preceding sound, is higher than those for sound fields C and G, which consist of a preceding sound and one following sound (C) or two following sounds (G). Moreover, as in Experiment 1, adding another following sound decreased the word intelligibility score. However, the effect of appearance order was not significant for any sound field.

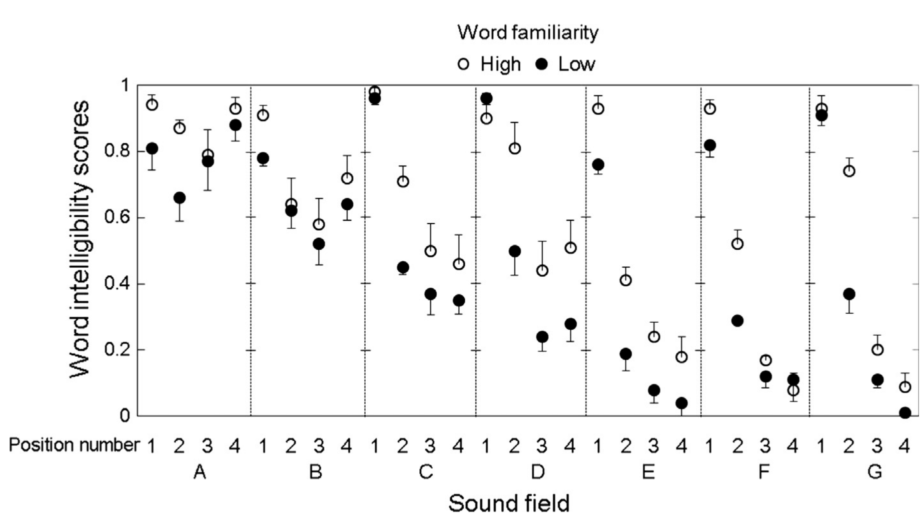


Fig. 4. Word intelligibility scores for each word position in a quadruplet for each sound field. A symbol shows a score averaged over seven listeners.

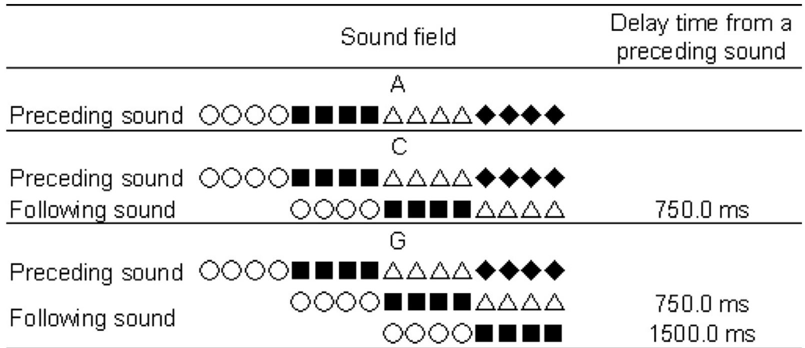


Fig. 5. Pattern diagram of the three sound fields used in Experiment 2. A cluster of the same four symbols (circles, squares, triangles, or rhomboids) represents an individual four-mora word, one symbol represents each mora of a word. Open and filled symbols respectively represent high- and low-familiarity words.

Table 2
Assignment of listeners to speech unit/sound field combinations.

Speech unit (high-low order)	Sound field			Speech unit (low–high order)	Sound field		
	A	B	C		A	B	C
I	L1	L2	L3	I′	L4	L5	L6
II	L3	L1	L2	II′	L6	L4	L5
III	L2	L3	L1	III′	L5	L6	L4
IV	L1	L2	L3	IV′	L4	L5	L6
V	L3	L1	L2	V′	L6	L4	L5
VI	L2	L3	L1	VI′	L5	L6	L4
VII	L1	L2	L3	VII′	L4	L5	L6
VIII	L3	L1	L2	VIII′	L6	L4	L5
IX	L2	L3	L1	IX′	L5	L6	L4
X	L1	L2	L3	X′	L4	L5	L6

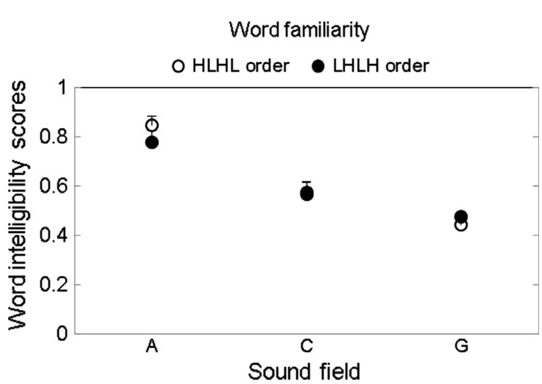


Fig. 6. Word intelligibility of sound fields of HLHL order and LHLH order.

Two-way ANOVA was carried out with appearance order and sound field as factors. The result reveals the significant effect of sound fields ($F(2,20) = 101.15$, $p < .001$). Multiple comparisons on the main effect of sound fields reveal that the word intelligibility of the condition with a preceding sound only (sound field A) is significantly higher than that of the sound fields with following sounds (sound fields C and G), and that the word intelligibility scores of sound field C (one following sound) was higher than that of sound field G (two following sounds). The effect of appearance order was not significant.

It is considered whether or not word familiarity affects word intelligibility when high- and low-familiarity words coexist in one sentence. Fig. 7 shows the word intelligibility of preceding sounds obtained for each word position, and for each of the familiarities of the preceding sound (high or low), by pooling the appearance order. In Fig. 7, the color of the circle indicates the familiarity of the target four-mora word at the target position of the quadruplet. Because different familiarity words appeared alternatively in the experiment, the quadruplet used to calculate the intelligibility scores of the first and the third words was different from that used to calculate the intelligibility scores of the second and the fourth words under the same experimental conditions. For almost all word positions except for the second and fourth word in sound field A, where no following sound was presented, the intelligibility of high-familiarity words is higher than that of low-familiarity words. Three-way ANOVA was carried out using word familiarity, sound field, and word position as factors. The result reveals the significant effects of word familiarity ($F(1,10) = 9.96$, $p < .05$) while the other two factors were not statistically significant. This means that the intelligibility of high-familiarity words is higher than that of low-familiarity words under long-path echo conditions, and when high- and low-familiarity words coexist in one sentence.

4. General discussion: Effects of overlapping high- and low-familiarity words on word intelligibility

In the previous two sections, the effect of the preceding sound's word familiarity on word intelligibility was discussed, but the effects of overlapping high- and low-familiarity words were ignored. The results showed that the intelligibility of high-familiarity words is higher than that of low-familiarity words. However, it is impractical to compose announcements that contain only high-familiarity words. In actual situations, high-familiarity words are occasionally overlapped by low-familiarity words that follow, and vice versa. The intelligibility of high-familiarity words (target) might depend on the familiarity of the word that overlaps the target word. If the intelligibility score of high-familiarity words would decrease more when overlapped by other high-familiarity words than when overlapped by low-family words, the role of word familiarity might be less important when optimizing speech

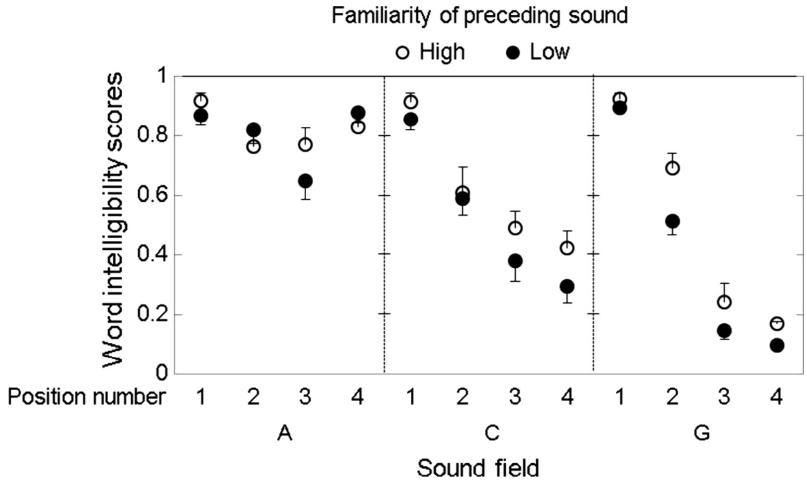


Fig. 7. Word intelligibility of preceding sound obtained for each word position of each sound field in Experiment 2.

intelligibility in the presence of long-path echoes. Here, we discuss the effects that overlapping high- and low familiarity words have on word intelligibility, based on the results for sound field C (which represents a simple condition common to the two experiments).

For each word position, Fig. 8 shows the word intelligibility of the target words in the four overlap conditions (H/H, H/L, L/H and L/L), in sound field C in Experiment 1 and Experiment 2. For example, H/L indicates the condition in which a preceding high-familiarity sound is overlapped by a following low-familiarity sound. In this figure, H of H/L and L of L/H represent the target words, while L of H/L and H of L/H represent the overlapped words. Scores for position 1 are not shown because it was not overlapped by a following sound in sound field C. It is clearly shown that, for all word positions, the intelligibility of high-familiarity words is high even when they are overlapped by other high-familiarity words.

Two-way ANOVA was carried out using overlap condition and word position as factors. The result revealed the significant effects of both overlap condition ($F(3,54) = 4.31$, $p < .01$) and word position ($F(2,54) = 11.25$, $p < .001$). Multiple comparisons on the main effect of the overlap condition showed no significant difference between H/H and H/L conditions, but the scores in H/H were significantly higher ($\eta^2 = 0.24$, $\beta = 0.03$) than those in L/H and L/L.

The fact that no significant difference was shown between the H/H and H/L conditions means that the intelligibility of high-familiarity words is not significantly affected by the familiarity of the overlapping word. Moreover, intelligibility in the H/H condition is significantly higher than that in the L/H condition. This means that intelligibility decreases to a greater extent when a preceding word with low familiarity is followed by a high-familiarity word than when both the preceding and following words have high familiarity. This shows the importance of using high-familiarity words. In other words, announcements presented from mass notification sound systems should be composed of high-familiarity words as much as possible.

We now consider in more detail the effect of long-path echoes on speech understanding. In the presence of a long-path echo, people listen to not only the direct sound but also the long-path echo sound. This means that people have two opportunities to hear the target sound, although some parts of the sound are overlapped. Such listening conditions are completely different from the situations in which people listen to target sounds degraded by various noises. Unlike such noisy conditions, it is important to segregate the target sound stream from other streams when people hear a sound in the presence of long-path echoes. Numerous studies have

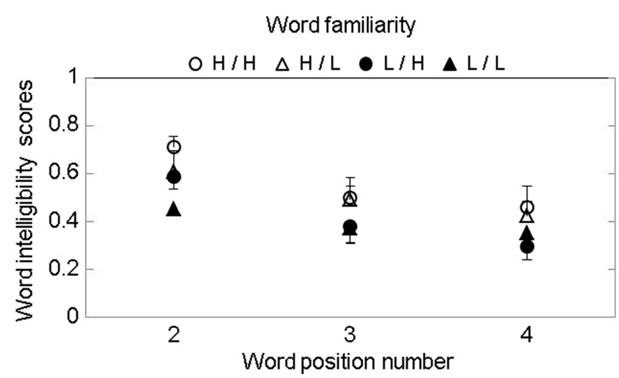


Fig. 8. For each word position, the word intelligibility in the four overlap conditions (H/H, H/L, L/H and L/L), in sound field C in Experiment 1 and Experiment 2. H of H/L and L of L/H represent the target words, while L of H/L and H of L/H represent the overlapped words.

pointed out that high-familiarity words can be robustly heard under various noisy conditions [13]. Considering this knowledge and our experimental results, the intelligibility scores of high-familiarity target words may have decreased more when overlapping another high-familiarity word than when overlapping a low-familiarity word. However, the results of this study clearly indicate that high-familiarity words are easily understood even when overlapping other high-familiarity words. This may be the case because high-familiarity words are easily segregated from other acoustic distractors (including other high-familiarity words), resulting in high intelligibility even when they overlap other high-familiarity words.

We consider the results obtained in the present study to be useful in improving speech intelligibility of the messages presented via mass-notification sound systems, and we did not investigate the effect of very low familiarity words. This is because the present study clearly shows the positive effect of the use of high-familiarity words in the presence of long-path echoes. However, very low familiarity words might be presented, very rarely though, such as place names for strangers or specific technical words to explain emergency. It may be necessary to confirm whether the tendencies found in the present study are also valid for low familiarity words. We are considering this as a future study.

While only female speaker's voices were used as the stimuli in these experiments, the absolute value of the intelligibility score might be affected by gender. However, Amano et al. clearly indicated that listeners can hear the high-familiarity words robustly under various signal-to-noise ratio (SNR) conditions regardless of the difference of the talkers including gender [13]. Therefore, the tendency of the intelligibility scores obtained using a female speaker's voice would be identical to those obtained when a male speaker's voice was used as the stimulus.

5. Conclusions

This research investigated the effects of word familiarity on the word intelligibility of four continuous words under long-path echo conditions. The results of Experiment 1 clearly indicate that the intelligibility of high-familiarity words is higher than that of low-familiarity words under long-path echo conditions, regardless of the word position. The results of Experiment 2 demonstrated that the intelligibility of high-familiarity words is higher than that of low-familiarity words under long-path echo conditions, even when high- and low-familiarity words coexist in one sentence. These facts suggest that high-familiarity words are more robust against the influence of long-path echoes than low-familiarity words. Therefore, announcements presented from mass notification sound systems should consist of high-familiarity words as much as possible.

Acknowledgements

Authors wish to thank Dr. Jorge Trevino for his advise in preparing this paper. A part of this study was supported as a research and development project (Research and development strengthening disaster-proof properties of information and communication networks) by the 3rd supplementary budget of the Ministry of Internal Affairs and Communications, Japan, the A3 Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet," and Grant-in-Aid for Scientific Research (B) (No. 16KT0100).

References

[1] The Radio Use Web Site. <http://www.tele.soumu.go.jp/e/index.htm>.

[2] Toida Y. Sentence intelligibility of reinforced speech under echoic and disturbing noise conditions. Trans Archi Inst Jpn 1984;346:112–23 [in Japanese].

[3] Toida Y. Speech intelligibility in open air sound fields. J Acoust Soc Jpn 1987;43:519–25 [in Japanese].

[4] Toida Y. Characteristics and prediction of speech intelligibility in reinforced sound fields disturbed by echoes with a long delay time. J Archi Plan Env Eng 1997;492:1997 [in Japanese].

[5] Sato Ha, Morimoto M, Sakamoto S, Suzuki Y. The effect of long-path echo on intelligibility of the disaster radio system. Proc Spring Meeting Acoust Soc Jpn 2013:949–52 [in Japanese].

[6] Amano S, Kondo T. Lexical properties of Japanese, vol. 1. Tokyo: Sanseido; 1999 [in Japanese].

[7] Sakamoto S, Suzuki Y, Amano S, Ozawa K, Kondo, Sone T. New lists for word intelligibility test based on word familiarity and phonetic balance. J Acoust Soc Jpn 1998;54:842–9 [in Japanese].

[8] Sato Hi, Sato Ha, Yoshino H, Suzuki Y, Amano S, Kondo T, et al. Effects of word familiarity and hearing loss with aging on word intelligibility in noise and reverberation. J Acoust Soc Jpn 2002;58:346–54 [in Japanese].

[9] Morimoto M, Sato Hi, Kobayashi M. Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces. J Acoust Soc Am 2004;116:1607–13.

[10] Speech Resources Consortium. <http://research.nii.ac.jp/src/FW07.html>.

[11] Zhenglie Cui, Shuichi Sakamoto, Yuki Tamura, Yôiti Suzuki. Speech intelligibility under long-path echoes by using four-continuous-words simulating emergency mass-notification messages. In: Proc of the auditory research meeting, EA2016-26; 2016. p. 379–83 [in Japanese].

[12] Kohei Kinoshita, Masayuki Morimoto, Hayato Sato, The effect of a single long path echo and word familiarity on word and sentence intelligibility. In: Proc ASJ Autumn Meeting, 3-8-5; 2012. p. 1235–8 [in Japanese].

[13] Amano S, Sakamoto S, Kondo T, Suzuki Y. Development of familiarity-controlled word lists 2003 (FW03) to assess spoken word intelligibility. Speech Commun 2009;51(1):76–82.