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
Available online 3 April 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
0003-682X/

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.