Effective announcements during disasters for immediate evacuation: Emphasis on each sentence
Sayoko Takano* and Yoshio Tsuchida
Kanazawa Institute of Technology, 7–1 Ohgigaoka, Nonoichi, 921–8501 Japan
(Received 2 July 2020, Accepted for publication 5 October 2020)

Keywords: Disaster prevention broadcast, Announcements, Speech synthesis, Evaluation

1. Introduction
The 2011 Tōhoku earthquake and the tsunami in Japan reportedly caused more than 15,000 deaths. Most of the deaths were caused by the tsunami rather than directly by the earthquake or fire. Some of those affected may have listened to the announcements though radio speaker for immediate evacuation and took necessary action; however, others who did not were unable to save their lives.

In our research group, we investigated effective warnings for evacuation announcements by training the speakers. In the system, the speakers listen to the simulated surrounding noise and the long path echo, and try to overcome the noise of adverse weather conditions for producing a clearer effect [1]. Instructions are then given to be “louder,” sound as if in “slight panic,” adopt a “commanding” tone, and so on. Professional speakers changed their speech effectively and acquired a better score on residents’ “intention to evacuate.” On the other hand, non-professionals could not change their voice effectively. We investigated the voice parameters of these samples. The professionals changed the height, speed, and amplitude of their speech, while the non-professionals were unable to do so [2,3].

Effective warning announcements were examined by changing the various kinds of speech parameters showing that urgency can be perceptually increased by changing those parameters [4]. We also used similar techniques by changing the speech parameters focusing on the sentence meaning of the three sentences. A non-professional’s “normal” speech was changed in terms of its fundamental frequency (F0), speech rate using STRAIGHT [5]. This software is a speech transformation tool by analyzing mainly fundamental frequency and spectrogram, and enables to re-synthesize natural speech sound.

Subjects judged the expressions by listening to the re-synthesized speech. The better announcements with various emphasis were tested aiming at effective communication during disasters for quick evacuation.

2. Experiment
2.1. Speech materials
In our preliminary experiments for the better announcements during disaster prevention, we have acquired speech samples from two professionals and 6 non-professionals [2,3].

In this experiment, the speech sample of one male and one female were selected; their articulation was found to be better than that of the rest. The announcements in the speech samples were (1) “Chikaku no kawa ga harran shimasita” (A river in the neighborhood is in flood). (2) “Taihen kikenna jouai desu” (It is very dangerous). (3) “Tadachini shiteino syougakkouni hinan shitekusadasai” (Please evacuate to the closest elementary school immediately). We named (1) as “fact,” (2) as “status,” and (3) as “command” based on the meaning of the sentences.

2.2. Speech synthesis
The emphasized speech sounds were created using STRAIGHT [5]. The combination of amplitude (A), fundamental frequency (F), and speech rate (R) were manipulated. When emphasizing fact, for example, seven kinds of speech sounds were created, i.e., for three single (A, F, R), three dual (A-R, A-F, F-R), and one triple deformation (A-F-R). The same manipulation was also applied for the “status” and the “command.” One speakers’ natural speech was manipulated, and 21 types of sentences were formed. In this experiment, speech by one male and one female was used.

The manipulations were set at natural speech. The amplitude was changed by multiplying 0.4 to the other parts, meaning, the emphasis was realized by making that part relatively louder than the other parts. The fundamental frequency was 1.2 times the original, and the speech rate was 0.85. The reasons for the selections of these parameters were the largest amount that the listeners feel natural even with the change.

2.3. Hearing experiment
Ten male adults with normal hearing were the subjects of the experiment. They listened to the speech sounds on loud speakers (BOSE, 101MM) in an un-echoic chamber. Additional noisy noise and long path echo were added using an acoustic simulator of a broadcasting system for disaster prevention (AEGIS, ver. H1A3). AEGIS is a customized software made by Mr. Koyama on music sound graphical programming interface (MAX7, cycling ’74). The maximum loudness was set to 73 dB.

The speech sounds were played randomly. The subjects were asked to evaluate “Nigerukini naru (eager to escape),” “Kikikan ga moteru (emergent),” “Jyudaini kanjuru (important),” and “Shinraiga moteru (trustworthy).”

The “natural” speech was set as the standard stimulus with an evaluation score 0, and the manipulated sound was the target. The subjects were asked to evaluate the manipulated

*e-mail: tsayoko@neptune.kanazawa-it.ac.jp
[doi:10.1250/ast.42.200]
3. Results and discussion

3.1. Emphasized parts

The average value of the scores from the hearing experiment conducted on 10 subjects is described in Fig. 1. The emphasis on “fact” (top) shows a higher value for “importance.” The emphasis on “status” (middle) shows a higher value for “emergence.” The emphasis on “command” (bottom) shows a higher value for “eagerness to escape” and “importance.”

Thus, emphasis was found to mainly affect the contents of the sentence.

3.2. Emphasis method

Effective emphases were investigated using analysis of variance (ANOVA, 5%). The effective emphasis for both men and women subjects were found to be (1) faster speech rate for the first sentence (fact), (2) higher fundamental frequency for the second (status), and (3) a higher fundamental frequency for the third sentence (command).

A higher fundamental frequency does not produce any effect if used in the first sentence (fact) because emphasis of a higher frequency is not effective when applied in a leading sentence while it becomes effective only when there is a preceding sentence spoken at a lower fundamental frequency.

4. Conclusion

A non-professional speaker may be required to broadcast evacuation announcement when disaster approaches, and in such cases, an effective emphasis would be to speak the first sentence (fact) faster, and the second (statement) and third sentence (command) at a higher fundamental frequency.

Acknowledgment

This work has been undertaken by students: Takato Hamada, Mitsuharu Ishii, and Kenichi Kojima, who graduated in 2017. We would like to thank Editage (www.editage.com) for English language editing.

This research is supported by JSPS 18H03490 and JST (JPMJMI18D1).

References