

公益財団法人 セコム科学技術振興財団
研究成果報告書

研究課題名

「災害時に必要な情報を音声により確実に伝える」
インテリジェント避難誘導音声呈示システムの研究開発

Study on intelligent voice presentation system for evacuation guidance

研究期間

平成 27 年 4 月 1 日 ～ 平成 31 年 3 月 31 日

報告年月

令和元年 6 月

研究代表者

北陸先端科学技術大学院大学 教授

赤木 正人

Professor, Graduate School of Advanced Science and Technology,

Japan Advanced Institute of Science and Technology

Masato AKAGI

Abstract

Japan is a disaster-prone country. There are many earthquakes, and in recent years, damages from wind and rain has also increased. It is important from a point of view of minimizing the damages to present appropriate voice evacuation guidance to a safe place according to the situation such as life related emergency.

Evacuation guidance by speech is used in various places. This is because it is effective even if induction signs cannot be confirmed visually, and many people can be given guidance instructions at one time. However, on the other hand, it is also a fact that the evacuation guidance voices are difficult to understand and difficult to hear. The evacuation guidance voices in environments with a large amount of noise and a long reverberation sometimes cause a problem that the voice is difficult to hear. Simply increasing the volume cannot improve the hearing audibility of the voice. Also, even if it can be heard, it may not recognize danger due to "Normalcy Bias", and as a result, evacuation may be delayed. Evacuation guidance that strongly conveys the danger that exists there and strongly urges evacuation is necessary, that is, speech announcements according to various environmental factors and the degree of danger.

Our research group have studied for the purpose of proposing and constructing a speech synthesis/conversion system that can perform evacuation guidance with high intelligibility in order to "provide necessary information by speech." In this report, in order to achieve this purpose, we introduce the research contents that our group has conducted about the methods of appropriately presenting evacuation guidance speech.

When presenting evacuation guidance, it is necessary to use appropriate speech to hear the speech intelligibly even in noisy and reverberant environments and to recognize how urgent the evacuation request is. In the conventional method, since evacuation guidance was presented unilaterally with recorded (or synthesized) speech regardless of the sound environment, it was unclear whether all the presented voices were properly heard. In the proposed method, in order to realize a speech synthesis/conversion system capable of evacuation guidance with a high degree of intelligibility, the following two tasks were set. (1) Adaptive control of presented voice by feedback of sound environment: By constantly monitoring the presented speech while measuring the noisy reverberant environment with multiple microphones, generate announcement speech to the evacuation guidance site.

The presented speech is as natural as Lombard effect and is the most intelligible. (2) Control of linguistic and paralinguistic information according to the situation: Generate announcement speech by selecting linguistic information and adaptively adding paralinguistic information according to the situation so that the listeners can be alerted. In speech synthesis/conversion, by considering what properties evacuation guidance speech should have from knowledge of emergency speech perception, we generate highly natural speech based on the restriction of human speech production mechanism.

As a result of the study, about task (1), it has been confirmed that the Lombard speech simulated by our system is intelligible and natural as almost the same as that of the Lombard speech spoken by humans. From now on, it is necessary to synthesize more intelligible speech that can surpass human's Lombard speech while maintaining the naturalness of synthesized/converted speech. Even in reverberant environments, we were able to find voices that can improve intelligibility during reverberation. In the future, it is necessary to explore rules for simulating this speech and apply it to speech synthesis. About task (2), we analyzed perception and acoustic features of the voices that were uttered by professional TV news casters in a tsunami disaster, using urgency as a keyword. It was revealed that the fundamental frequency is most relevant to the perception of urgency. It is a future work how the system controls speech with urgency according to the situation by the rule.