

Title	振幅包絡線情報の局部時間反転による音声プライバシー保護の研究
Author(s)	坂本, 貴望
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17098
Rights	
Description	Supervisor: 鷗木 祐史, 先端科学技術研究科, 修士 (情報科学)

Study on speech privacy protection with locally time-reversing temporal amplitude envelope

1910100 SAKAMOTO Takami

Speech is one of the most important tools of communication in our social life. Speech contains three types of information: linguistic information, non-linguistic information, and para-linguistic information. Linguistic information is a verbal message that can be expressed in language and written. Non-linguistic information relates to the speaker's age, gender, personality, emotions, and so on. In general, these information cannot be intentionally controlled by the speaker. Para-linguistic information includes accent, intonation, speaking speed, and voice pitch, all of which cannot be expressed in writing. These information are intentionally added by the speaker to transform the linguistic information. In particular, linguistic information, which includes the semantic content of a message, is important in speech communication. It is not a problem for people who are communicating with each other to talk about linguistic information. However, there are situations in which it is undesirable for linguistic information to be overheard by unintended listeners, such as when talk about private conversations. In situations where confidentiality is required, linguistic information needs to be protected strictly and appropriately.

In open spaces such as hospitals, pharmacies, and conference rooms, people often talk about personal or confidential information. If such private conversations are overheard by unintended listeners, personal or confidential information may be leaked. Problems arising from private conversations must be solved to protect speech privacy.

Akagi and Irie's research on speech privacy protection focuses on mishearing of speech. This method is based on Acceleration of perceptual fusion. Acceleration of perceptual fusion is a phenomenon in which two or more different sounds are perceived as a single sound. It is occurred according to Bregman's psychoacoustical heuristic regularities. This regularities composed of the following four rules: (1) common onset or offset, (2) smoothness, (3) harmonicity, (4) common changes occurring in the acoustic events. In this research, the target speech (conversational speech) and sound for hearing protection are presented simultaneously. It is shown that the two sounds are heard as one, and the speech content is obscured. While this method is very effective as a method of speech privacy protection, it has the problem of causing discomfort due to excessive deformation of the spectral envelope.

In auditory perception, the temporal structure of speech sounds has a significant effect on listening comprehension. For example, in Drullman's

research, it is investigated whether the temporal amplitude envelope(TAE) or the temporal fine structure(TFS) of a sound contains more information related to speech intelligibility. The results showed that TAE plays an important role. In addition, in Ueda's research is showed that speech intelligibility is significantly reduced by locally time-reversing. Locally time-reversing is a method of dividing speech into short segments, reversing the time within each segment, and then connecting the segments. It has been shown that when the locally time-reversal length is short (20-40 ms), the speech can be understood, but as the locally time-reversal length increases, the speech cannot be understood.

These results suggest that an effective method of speech privacy protection can be achieved by locally time-reversing TAE. This paper aims to achieve an effective method of speech privacy protection by directly processing spoken language information in time domain. In this paper, the following three points are focused on. (1) TAE plays an important role in speech intelligibility. (2) target speech and locally time-reversed speech have the same TFS based on the most restrictive of Bregman's four psychoacoustical heuristic regularities, rule(3): harmonicity. (3) locally time-reversing is used to manipulate the temporal structure of speech. From previous study, it has been found that two speech with different TAE (one of them locally time-reversed TAE) and the same TFS are most perceptually fused than those with different conditions from this. Therefore, in this paper, it is investigated whether locally time-reversing TAE can be perceptually fused to the target speech and reducing the speech intelligibility of the target speech under similar conditions.

In this study, two experiments were conducted. The first experiment was conducted in a soundproof room using headphones. The second experiment was conducted in an audio-visual laboratory using speakers. In the first experiment, it is considered whether or not that intelligibility of a target speech can be reduced with locally time-reversing TAE while TFS of the target speech has not been manipulated. Word intelligibility test with familiarity-controlled word lists was conducted to clarify these two points under two conditions of the highest and lowest word-familiarity levels and under nine conditions of locally time-reversing length (20, 40, 80, 160, 240, 320, 480, 640 ms, and whole duration). Results were summarized as: (1) the word recognition rate is reduced by controlling locally time-reversal length. (2) reduction degree of the word recognition rate depends on word-familiarity levels, that is, 77% at the highest familiarity level and 44% at the lowest familiarity level. (3) the most effective length of locally time-reversal is around 160 ms. These findings indicate that locally time-reversing TAE enables to reduce speech intelligibility of the target speech.

In the second experiment, the effectiveness of the method in a real environment is confirmed. In this test, the conditions of word-familiarity levels and locally time-reversing length were the same as in the previous test, and the target speech and locally time-reversing speech were played from different speakers. The results showed that locally time-reversing TAE enables to reduce speech intelligibility of the target speech in a real environment. It is also found that the effect is greater than that of the previous test that using headphones.

In both experiments, it was possible to reduce speech intelligibility of the target speech. These results suggest that locally time-reversing TAE enables to reduce speech intelligibility of the target speech. Also, it is shown that this method can effectively protect speech privacy.