

Intelligibility of English mosaic speech: Influence of manipulating mosaic block duration

サンティ

<https://hdl.handle.net/2324/4475136>

出版情報 : Kyushu University, 2020, 博士 (芸術工学) , 課程博士
バージョン :
権利関係 :

Name: Santi

Dissertation title: Intelligibility of English mosaic speech: Influence of manipulating mosaic block duration
(英語モザイクスピーチの明瞭度：モザイクブロック幅の操作による影響)

Category: 甲

Abstract of Dissertation

Mosaic speech is degraded speech that is segmented into time \times frequency blocks. Earlier research with Japanese mosaic speech has shown that the intelligibility of mosaic speech was almost perfect for mosaic block durations (MBD) of 20 and 40 ms. The first objective of the present research was to investigate the intelligibility of English mosaic speech, and whether its intelligibility would vary if it was compressed, preserved, or stretched in time. The second objective was to investigate whether the effects of compressing, preserving or stretching mosaic speech would be similar among listeners with different language backgrounds. To achieve these objectives, two experiments were conducted. The preliminary experiment was conducted first with Indonesian listeners (n= 20) followed by the main experiment with native-English (n= 19), Indonesian (n= 19), and Chinese (n= 20) listeners. In the experiments, English mosaic words were presented to the participants, and they typed what they had heard. The intelligibility of English mosaic speech (individual words) was obtained by counting the number of correct words given by the participants.

For the first objective, from the two experiments conducted, it was found that the listeners from the three language groups (native-English, Indonesian, and Chinese) showed the same trends in intelligibility scores: English mosaic speech was most intelligible when the OMBDs were preserved or stretched into 20- or 40-ms MBDs. Intelligibility decreased when the OMBDs were compressed, or stretched into MBDs of 80 ms or longer. The results seem to agree with the results of earlier research, which showed that mosaic speech is most intelligible in the segment duration range of 20 to 40 ms. When the segment is longer, the intelligibility becomes lower, but the results thus show that even rather long 40-ms segments can be processed as intelligible speech.

The second objective of this thesis research was to investigate whether the effects of compressing, preserving, or stretching mosaic speech varies among listeners with different language backgrounds. The results showed that the intelligibility was relatively high for stimuli with preserved

OMBDS of 20 ms and 40 ms for all language groups, and also for stimuli with an MBD of 40 ms after stretching the OMBD of 20 ms, but only for the native-English group. The OMBD was manipulated by compressing or stretching it without changing its linguistic information. However, the speed of speech changed and this caused the intelligibility to change as well. Both non-native listener groups showed the same trend regarding the speed of speech, that is, the intelligibility was highest for the preserved speech. However, the native-English listeners obtained the highest intelligibility scores for preserved speech or slightly slower speech, but this happened only when the OMBD of 20 ms was stretched into a 40-ms MBD. Thus, this thesis research suggests that presenting the same acoustic information in any temporal segment does not guarantee that the intelligibility will be preserved, but the temporal segment duration plays the most important role to determine intelligibility in mosaic speech perception. In other words, the intelligibility was affected by mosaic block duration (MBD). Regarding the stimuli with the same (preserved/stretched) MBDs among both OMBDs, i.e. 40, 80, or 160 ms, the intelligibility did not change significantly even when the amount of information must have changed for the native-English and the Indonesian listeners, but not for the Chinese listeners. Thus, this thesis research suggests that the intelligibility was not affected by OMBD when it was preserved/stretched in the range of 40-160 ms.

In general, the results of the thesis research suggest that humans can extract new information from individual speech segments of about 40 ms, but that there is a limit to the amount of linguistic information that can be conveyed within a block of about 40 ms or below.