Calculation of speech intelligibility using four orthogonal factors
extracted from the autocorrelation functions of source and sound field signals

Tetsuichi Shoda and Yoichi Ando

Graduate School of Science and Technology, Kobe University, 1 Rokkodai, Nada, Kobe 657 8501, Japan

Abstract: In this paper, a method of calculating speech intelligibility is proposed. This method is based on the model of auditory-brain system including the autocorrelation (ACF) mechanisms. The distance between the syllables calculated by use of the four factors extracted from the ACF of source and sound field signals. Using these distances and its linear combination, speech intelligibility may be calculated. Results of speech intelligibility tests as a function of the delay time of the single reflection are in good agreement with calculated.

INTRODUCTION

Acoustical index for evaluating speech intelligibility in an acoustic space, such as speech transmission index (STI) are widely applied. However these indexes are calculated only by the effect of sound field. In order to calculate speech intelligibility of single syllables, we should consider the characteristic of source signal and the effect of temporal structures of sound fields. In this study, a new method to evaluate speech intelligibility reflecting these effect is proposed. This method is based on the ACF mechanism of the speech information processing assumed in the auditory pathway and the brain.

THE MODEL OF CALCULATING SPEECH INTELLIGIBILITY

Supposing the existence of distance measure between single syllables, speech intelligibility may be calculated from a simple equation. Let us assume that there is a characterization of each isolated template syllable $S^T_k$ in the brain, in the form of $C^T_k$, where $T$ means the template of the syllable. And let $S^F_k$ be a test utterance produced in an sound field, where $SF$ means the sound field. The distance between $S^F_k$ and $S^T_k$ can be given by

$$d_k = D(S^F_k, S^T_k) = |C^F_k - C^T_k|$$

Let the number of syllables of the same category with $S^F_k$ be $N$. (syllables can be categorized by the type of the consonant and the vowels, because listeners don't mistake syllable over the category.) The equation for calculating intelligibility of the syllable $S_k$, $\Psi_k$ may be obtained

$$\Psi_k = 100 \cdot N \cdot \exp \left( \frac{d_k}{d_1} \right)$$

CHARACTERIZATION FOR CALCULATING DISTANCE

In modelling the auditory pathway and brain, the power density spectra in the neural activities in auditory pathways are transformed into the ACF. If $\tau_{ACF} = 1$, the speech signal of the ear entrance may be processed as the ACF (temporal criterion) in the left hemisphere. Four orthogonal factors are extracted from the ACF to characterize single syllables. As shown in FIGURE 1, the orthogonal factors are, [1] $\tau_e$: the effective duration of ACF (the ten percentile delay) meaning the repetitive feature of signals; [2] $\tau_1$: the delay time of the first peak of ACF signifying the pitch of signals; [3] $\phi_1$: the amplitude of the first peak indicating the strength of the pitch; and [4] $\Phi(0)$: the power of the signal frame.
CALCULATION OF SPEECH INTELLIGIBILITY

In this study the intelligibility of the sound field which consists of the direct sound and the single reflection was calculated. The amplitude of the reflection was same as that of direct sound, and the delay time of the reflection, $\Delta t_1$, was varied from 0 to 480 ms. The test signal is consisted of 50 Japanese monosyllables placed between maskers. The direct sounds without maskers are used as templates. Then, the ACFs of the template and the test syllables were analyzed. Actually, two types of distances were calculated with the four factors, one is the distance between the template and the direct sound, and the other is that between the template and the reflection. Shorter results of two types of distances was adopted in calculating speech intelligibility. Using this distance and the equation (2), intelligibility are calculated by each factors. Combining them linearly by multiple regression analysis, the intelligibility of all syllables were calculated.

CONCLUSION

Result of calculated and experimental intelligibility are shown in FIGURE 2. It is remarkable that the calculated value are in good agreement with the experimental one. Therefore it is concluded that the four factors extracted from the ACF of thr source signals and thr sound field signals may play an important role to recognize speech.

REFERENCES