学位報告 4

報告番号	※甲　　第　　号

主　論　文　の　要　旨

論文題目 Speech characteristics during conversation and their importance on information transmission

（対話音声における話者特徴とその情報伝達における重要性について）

氏　　名 陳　伯翰

論 文 内 容 の 要 旨

Language production and comprehension have been studied by scientists for decades, And hundreds of empirical observations as well as experimental results have been reported. Numerous models and theories have been proposed to describe a wide ranging aspects of human communication and its underlying mechanisms. This dissertation will investigate the speaker-specific speech characteristics and their impact on information transmission. After reviewing related researches, I consider that similarity in the speaker-specific speech characteristics of conversation partners is one of the key to understanding high linguistic information transmission efficiency during conversations. Since few systematic investigations of the role of speaker-specific speech characteristic similarity in conversational information transmission have been published, especially ones focus on segmental speech characteristics similarity, the objectives of this dissertation are as follows:

1, Investigate the effective similarity measure of speaker-specific speech characteristics on speaker diarization tasks while developing speaker diarization system with higher clustering accuracy (first study);

2, Investigate the impact of similarity in speaker-specific speech characteristics, especially the segmental-based speech characteristics, on subtle prosodic information transmission (second study);

3, Use similarity in the speaker-specific speech characteristics of interlocutors as a predictor of information transmission quality, and explore the effects of this

similarity on task performance using spontaneous speech corpus (third study).

In the background survey, segmental-based speech characteristics, prosodic speech characteristics and lexical speech characteristics were found to be the most important speech characteristics affecting performance in speaker related signal processing tasks (e.g., speaker identification, speaker recognition, speaker diarization, etc.).

In the first study, a method to evaluate whether a cluster contains all of one (and only one) speaker's speech segments, based on the statistical properties of with-cluster similarity scores and between-cluster similarity scores in a ``speaker space'' was proposed. Modified bottom-up clustering was then conducted based on the proposed cluster evaluation method in order to increase diarization accuracy by preventing over-merging. Experimental results showed that the proposed method achieve higher clustering accuracy than conventional bottom-up clustering methods.

In the second study, a listening experiment was designed to investigate the efficiency of subtle prosodic information transmission at different levels of speech characteristic similarity. Japanese right-branching (RB) vs. left-branching (LB) ambiguous sentences were used as experimental material. Morphing technology and text-dependent objective similarity measures were introduced to control similarity levels. Participants were asked to finish a target identification task with RB vs. LB materials as targets. Participant response time during the target identification task, as well as the proportion of eye-fixing on different targets were recorded for analysis. Results showed that speech characteristics similarity apparently has facilitative effect on prosodic information transmission.

In the third study, the impact of similarity in speaker and listener speech characteristics on the quality of linguistic information transmission was investigated, using a map task dialogue corpus. Similarity between the segmental (MFCC) voice features, prosodic features and lexical styles of different speakers were analyzed. Most of these similarity measures were shown to have significant facilitative effect on information transmission quality as measured with a direction following task in which a speaker is describing a route to a partner.

In general, experimental results showed that similarity in speaker-specific speech characteristics between conversation partners facilitated information transmission accuracy. The findings imply that self-similar voice is an effective direction towards high efficiency linguistic information transmission. The present dissertation is just the first step towards self-similar high efficiency information transmission system, several important issues, such as the impact of voice naturalness (caused by voice conversion) on linguistic information transmission and the comfortableness of hearing self-similar voice, are still unclear which need to be investigated in the future.