

# Specific Speaker's Japanese Speech Corpus over Long and Short Time Periods

Satoru Tsuge<sup>1</sup>, Masami Shishibori<sup>1</sup>, Fuji Ren<sup>1</sup>,  
Kenji Kita<sup>2</sup> and Shingo Kuroiwa<sup>1</sup>

<sup>1</sup> Faculty of Engineering, the University of Tokushima,  
2-1, Minami Josanjima, Tokushima, Japan  
tsuge@is.tokushima-u.ac.jp

<sup>2</sup> Center for Advanced Information Technology, the University of Tokushima,  
2-1, Minami Josanjima, Tokushima, Japan

**Abstract.** It is known that speech recognition performance varies pending when the utterance was uttered although speakers use a speaker-dependent speech recognition system. This implies that the speech varies even if a specific speaker utters a specific sentence. Hence, we investigate the speech variability of a specific speaker over short and long time periods for getting the stable speech recognition performances. For this investigation, we need a specific speaker's speech corpus which is recorded over long time periods. However, at present, we have not seen such a Japanese speech corpus. Therefore, we have been collecting the Japanese speech corpus for investigating the relationship between intra-speaker speech variability and speech recognition performance. In this paper, first, we introduce our speech corpus. Our corpus consists of six speakers' speech data. Each speaker read specific utterance sets three times a day, once a week. Using a specific female speaker's speech data in this corpus, we conduct speech recognition experiments for investigating the relationship between intra-speaker speech variability and speech recognition performance. Experimental results show that the variability of recognition performance over different days is larger than variability of recognition performance within a day.

## 1 Introduction

Recently, speech recognition systems, such as car navigation systems, and cellular phone systems have come into wide use. Although a speaker uses a speaker-dependent speech recognition system, it is known that speech recognition performance varies pending when the utterance was uttered. For this reason, we consider that speech characteristics varies even though the speaker and utterance remain constant. This intra-speaker variability is caused by some factors including emotion and background noise. If the recognition performance is not consistent, then products using speech recognition systems become less useful for the end-user. As the relationship between intra-speaker's speech variability and speech recognition performance is yet unclear, we began to investigate the nature of this relationship.