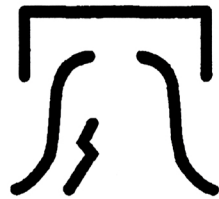# Proceedings

October 3-6, 1996
Wyndham Franklin Plaza Hotel
Philadelphia, PA, USA

ICSLP 96

Sponsored by

University of Delaware
Alfred I. duPont Institute

in conjunction with

Acoustical Society of America
Acoustical Society of Japan
American Speech-Language-Hearing Association
Australian Speech Science and Technology Association
European Speech Communication Association
IEEE Signal Processing Society
Incorporated Canadian Acoustical Association
International Phonetic Association
Linguistic Society of America

Fourth International Conference on Spoken Language Processing

3

CONSTRUCTING MULTI-LEVEL SPEECH DATABASE FOR SPONTANEOUS SPEECH PROCESSING

*Minsoo Hahn, **Sanghun Kim, **Jung-Chul Lee and ***Yong-Ju Lee

* Audio Information Processing Section
** Spoken Language Processing Section
Electronics and Telecommunication Research Institute
P.O. BOX 106 Yu-Seong Post Office, Taejeon, KOREA
Tel: 82-42-828-6234 Fax: 82-42-828-6231
e-mail: mshahn@audio.etri.re.kr
*** Dept. of Computer Eng.,WonKwang Univ., Chonbuk, Korea

ABSTRACT

This paper describes a new database, called muti-level speech database, for spontaneous speech processing. We designed the database to cover textual and acoustic variations from declarative speech to spontaneous speech. The database is composed of 5 categories which are, in the order of decreasing spontaneity, spontaneous speech, interview, simulated interview, declarative speech with context, and declarative speech without context. We collected total 112 sets from 23 subjects(male: 19, female: 4). Then the database was firstly transcribed using 15 transcription symbols according to our own transcription rules. Secondly, prosodic information will be added. The goal of this research is a comparative textual and prosodic analysis at each level, quantification of spontaneity of diversified speech database for dialogue speech synthesis and recognition. From the preliminary analysis of transcribed texts, the spontaneous speech has more corrections, repetitions, and pauses than the others as expected. In addition, average number of sentences per turn of spontaneous speech is greater than the others. From the above results, we can quantify the spontaneity of speech database.

1. INTRODUCTION

According to the advances of speech processing technologies, the spontaneous speech processing becomes one of the important topics. In particular, a speech synthesizer has to produce at least spontaneous speech-like output if it is adopted in spontaneous speech translation system.

The final goal of this work is to develop a dialogue type speech synthesizer with proper spontaneity by improv-ing our present text-to-speech system[1]. However the current text-to-speech systems mainly handle the declarative speech only, thus the knowledges acquired from the declarative speech processing cannot be directly applied to the dialogue speech synthesis system. And that is why many researchers started their studies on dialogue speech processing[2][3][4]. This paper can be considered as our first step for the dialogue speech processing.

This paper is organized as follows. The motivation of collecting multi-level speech database will be presented in section 2. In section 3, we will explain criteria to classify the multi-level speech. In section 4 and section 5, Collecting strategy and specification of multi-level speech database will be described respectively. Finally, we will present the preliminary results on multi-level speech database.

2. MOTIVATIONS

It is likely that various speech types will be synthesized before long[5]. The dialogue speech synthesis is more valuable to develop than the other speech type[6]. However, since only a few researches on dialogue speech processing have been performed, there are many problems to solve for implementing the dialogue speech synthesis system based on the current techniques.

First of all, the lack of knowledges about the characteristics of spontaneous speech makes more difficult to develop the natural dialogue type speech synthesis system. As you know, dialogue speech has various linguistic phenomena, i.e. interjections, anaphora, repetitions, and repairs. And also it has more dynamically prosodic changes than declarative speech. In speech recognition,