Chapter XIV
Synthetic Speech in Computer-Enhanced Foreign Language Learning

Min Kang
Kobe University, Japan

ABSTRACT

The goal of this chapter is to explain several experiments carried out by our research group to explore whether synthetic speech can be currently used to replace natural speech in listening materials for foreign language learning or not. For CALL purposes, synthetic speech in English was evaluated from the viewpoints of both foreign language learners and teachers. We conducted several surveys: (a) to find out if the synthetic speech generated by current TTS engines is as efficient as natural speech in training listening skills, (b) to identify the specific ways in which the evaluated synthetic speech is as good as natural speech, (c) to determine the relationship between changes in individual listening comprehension ability and the results of the quality evaluations of synthetic speech, and (d) to discuss the possible approaches for using synthetic speeches.

1. INTRODUCTION

The approach of computer-assisted language learning (CALL) has been widely recognized in foreign language education and is gradually being adopted in classroom learning. It has been indicated that computer-based training can be effective in improving a learner’s perception and production in target languages and that a CALL system can be designed to support collaborative learning (Chapelle, 1998; Hoven, 1999; Wang & Munro, 2004). CALL is a powerful means for teachers to provide students with a variety of learning styles. CALL systems make it possible not only to pres-
Synthetic Speech in Computer-Enhanced Foreign Language Learning

ent materials that can cater to individual needs in classrooms and carry out formative assessment, but also to increase opportunities for learners to receive listening input outside the classroom. It has been pointed out that exposure to the target language is very important in order to enhance listening comprehension ability (Hudson, 2000; Klein, 1986; Krashen, 1982). Unlike second language learners, foreign language learners have few opportunities for exposure outside the classroom. On the other hand, the CALL style has not necessarily achieved wide practical use. Although there are lots of CALL applications and e-learning tools to support this learning style, few of them have been utilized in classrooms as a regular tool. Developed CALL materials are costly and tend to be similar in content, which makes it difficult for teachers to adjust the materials according to students’ needs. Although there are numerous authoring tools, in addition to the learning content management systems (LCMS), which is designed to help teachers to create their own materials, an unavoidable aspect is the fact that managing listening materials costs teachers time and effort. The utilization of some tools may require advanced computer skills. The recording and editing of sound files continue to be the responsibility of teachers. In addition, the lack of a native speaker environment strongly affects the creation of listening materials. Therefore, for the wide use of CALL, it is crucial to develop applications with functions that efficiently help teachers to easily create their own listening material for the students’ need for exposure to the target language in training listening skills.

Speech synthesis technology may be helpful for improving the situation mentioned above and supporting the function. The speech synthesizer has significantly evolved since the release of MITalk-79 (Furui, 2002). The voice quality and intelligibility of current text-to-speech (TTS) systems have been improved significantly to the point that they are adequate for wide use in services and applications like voice response systems (Murray & Rohwer, 1996; Schroeter, 2006). Admittedly, TTS systems are not completely natural-sounding, but many voices are highly intelligible and natural-sounding, making it difficult to distinguish them from recordings of human voices. A great deal of research has focused on how to generate more natural sound output so that it is possible for machine-generated speech to approximate natural speech in intelligibility and naturalness in the near future. If a CALL system integrates a TTS engine, it will be possible for teachers to produce listening materials like reading materials. In fact, there are some CALL systems that have integrated TTS (ATRC, University of Toronto, 2008) and many electronic dictionaries have the integrated TTS systems.

Although speech synthesis technology is potentially beneficial to CALL, the use of synthetic speech has not been widely accepted. A possible reason may be that it is not yet clear whether synthetic speech is suitable and beneficial to language learning (Handly & Hamel, 2005). Because it is not completely natural-sounding, the use of synthetic speech as training materials for listening skills may cause a contrary effect if the suitability and benefits are not proven. For the past few decades, speech synthesis has been continuously used in exploring the mechanism of human speech perception. Meanwhile, examining TTS speech perception of listeners is an important stage in the development of TTS systems. With respect to TTS speech perception, a substantial amount of research has been conducted (Flanagan, 1972; Francis & Nusbaum, 2007; Logan, Lively & Pisoni, 1991; Schroeder, 1999). Some research also focused on training the language learner’s perception of synthetic speech but limited enhanced learners’ coding ability to specific several phonemes (Jamieson & Morosan, 1986; Strange & Dittmann, 1984). Despite a lot of effort being spent on speech perception, most CALL systems, especially those for classroom learning, have not integrated TTS engines. Why have so many remarkable results on TTS speech