

REKNEW-DYSARTHRIA

Abstracts from research conducted in collaboration with the OHSU Department of Engineering & Science, Division of Biomedical Computer Science and BioSpeech, Inc.

USER ADAPTATION OF AAC DEVICE VOICES - Phase I SBIR

Jan van Santen, PI and President of BioSpeech, Inc.

NIH 1R41DC008712-01

1/2007 - 6/2008

A wide range of individuals cannot communicate by voice. Voice enabled Augmentative and Alternative Communication (AAC) devices are often the only channel available by which these individuals can communicate. While many voice enabled AAC devices are currently available, they lack the important ability to generate customized speech that mimics aspects of the user's past or intermittently available speech. Modern "concatenative" speech synthesis technology can mimic a given speaker's voice, by excising speech fragments from a recorded speech data base ("acoustic inventory") and recombining these into output speech using sophisticated algorithms. It requires, however, a large amount of recordings and a high degree of consistency of pronunciation of the speaker. Many AAC users cannot meet these requirements because they already have lost the capability to speak or they cannot speak with adequate consistency of pronunciation. A new type of technology, voice transformation (VT) technology, is available that can transform speech spoken by a "source" speaker into speech that is perceived as spoken by a specific "target" speaker. To tune the transformation system, parallel "training recordings" of the same text are needed from the source and target speakers. The amount of training recordings is far less than what is needed for a high-quality acoustic inventory. We propose to use VT in combination with speech synthesis to convert the synthesis system's acoustic inventory into an acoustic inventory that mimics the target speaker's voice. The training recordings can consist of old home videos, or fragmented recordings produced during periods of intact speech, provided that they contain at least one sample of each phoneme. In Phase I, we will develop and evaluate a VT based synthesis system. The project will use high- quality and home-video quality recordings from male and female adults and children to create limited acoustic inventories (adequate to generate a specific set of test sentences) and VT training recordings. Perceptual experiments will be conducted to evaluate voice quality and perceived speaker identity. Phase II will focus on developing complete acoustic inventories for several canonical speakers that will be selected to cover a range of speaker characteristics, and on producing portable, user-friendly software. The anticipated commercial offering consists of (i) software components to be licensed to AAC vendors and (ii) a service consisting of collection and processing of recordings and creation of personalized acoustic inventories. Speech communication ability is impaired or absent in millions of Americans due to neurological disorders and diseases and to trauma, including autism, Parkinson's disease, and stroke. Augmentative and Alternative Communication (AAC) devices that are operated via switches, keyboards, and a

broad range of other input devices, and that have synthetic speech as output, are often the only manner in which these individuals can communicate. Without AAC devices, these individuals may suffer from severe social and psychological isolation, and may be unable to lead productive lives. A psychologically important feature that no currently available systems have is the ability to speak with the user's voice, i.e., the ability to produce speech that mimics the individual's pre-morbid speech or speech that the individual may be able to intermittently produce. The proposed project will use voice transformation (VT) technology to accomplish this goal. VT technology requires recordings of the user to be available, but there is substantial flexibility as to the nature and quantity of these recordings; they may consist of home videos or of fragmentary speech, provided that at least some samples are available of each speech sound in the language. The goal of the application is to develop a synthetic voice for an AAC system that sounds like the individual using the system (before they lost the ability to speak), without requiring very much recorded data on the part of the original talker. The system works by first creating a synthetic "base" voice (or set of base voices) using professional actors who must provide a fairly large inventory of speech data. Using the base voice and a small sample from the target talker (i.e., containing at least one instance of each phoneme), a new synthetic voice is created by essentially modulating parameters in the base voice so that it takes on characteristics of the target talker. The ability to create a voice that sounds like the original talker without much data from the original talker would be a significant advantage.