

CONTROLLO DELLA QUALITÀ VOCALICA PER LA SINTESI DEL PARLATO EMOTIVO

*Carlo Drioli, **Fabio Tesser, *Graziano Tisato, *Piero Cosi, *** Enrico Marchetto

* ISTC-SPFD CNR Istituto di Scienze e Tecnologie della Cognizione
Sezione di Fonetica e Dialettologia, Padova.

** Centro per la Ricerca Scientifica e Tecnologica, ITC-IRST, Trento.

*** Universita' di Padova, Dip. di Ingegneria dell'Informazione, Padova.

RIASSUNTO

La produzione del parlato in generale, e del parlato emotivo in particolare, e' caratterizzata da un'ampia varieta' di qualita' dell'emissione vocale, intesa come il risultato delle varie modalita' di fonazione possibili. Questo aspetto riveste un ruolo importante nella resa delle emozioni nella comunicazione verbale. E' comune osservare, infatti, in corpora di registrazioni di parlato emotivo, casi di fonazione non modale (voce morbida, sussurrata, laringalizzata, rauca, e cosi' via).

In questo lavoro si descrive un sistema di sintesi della voce che permette di controllare un set di parametri acustici di interesse per la simulazione di tipi di fonazione non modali. Il set di parametri di sintesi comprende, oltre ai tradizionali controlli di durata e pitch dei fonemi, controlli per l'intensita', l'enfasi spettrale, variazioni rapide di ampiezza e durata del periodo fondamentale di segmenti vocalizzati, variazioni lente di ampiezza e durata del periodo, spostamento delle formanti per distorsione dell'asse frequenziale, generazione di rumore di aspirazione.

Vengono proposte alcune regole di combinazione dei parametri acustici per la riproduzione delle seguenti qualita' di fonazione non modale: voce morbida (soft), alta (loud), aspirata (breathy), sussurrata (whispery), rauca (hoarse), tremante (tremulous). Viene discusso inoltre come queste diverse qualita' vengono utilizzate nella produzione del parlato emotivo.

Il sistema descritto si basa sull'ambiente di sviluppo per la sintesi vocale FESTIVAL e sulla tecnica di sintesi per concatenazione di difoni MBROLA. Viene fornita la possibilita' di includere nel testo da sintetizzare dei "tag" atti a specificare il tipo di fonazione o l'intenzione emotiva da utilizzare nella sintesi vocale. A questo scopo e' stato sviluppato per FESTIVAL il supporto per la gestione dei tag e per la conseguente generazione dei parametri di controllo per la sintesi attraverso appositi linguaggi di mark-up (APML/VSML).

Il controllo delle caratteristiche acustiche della sintesi si basa sull'elaborazione del segnale vocalizzato, prima della concatenazione tra difoni, mediante algoritmi specifici per i singoli parametri controllati. Vengono impiegate tecniche di elaborazione nel dominio del tempo per quelle trasformazioni che interessano il controllo del pitch, e tecniche nel dominio della frequenza basate su FFT e FFT inversa per quelle trasformazioni che interessano la struttura spettrale del segnale.

A conclusione della presentazione del metodo viene presentata una selezione di esempi percettivamente significativi con il fine di illustrare le caratteristiche e le potenzialita' del sistema presentato.