

MODELLIZZAZIONE DELLA PROSODIA E DEL TIMBRO PER LA SINTESI DEL PARLATO EMOTIVO

Mauro Nicolao, Carlo Drioli, Piero Cosi
Istituto di Scienze e Tecnologie della Cognizione - Sede di Padova "Fonetica e Dialettologia"
Consiglio Nazionale delle Ricerche
nicolao@pd.istc.cnr.it, cosi@pd.istc.cnr.it, drioli@pd.istc.cnr.it

SOMMARIO

Viene descritta una procedura per la creazione di una funzione di trasformazione di un segnale vocale neutro in uno caratterizzato emotivamente. Questa funzione è stata sviluppata sulla base di un modello statistico, a mistura di funzioni gaussiane, dello spettro del segnale vocale.

Sono utilizzati, come segnali di riferimento per l'addestramento del modello, due *database* di segnali vocali creati *ad hoc*: uno registrato da un parlatore, simulando l'emozione della collera, e uno neutro, con la stessa intonazione e durata dei fonemi, ottenuto con un sintetizzatore vocale per concatenazione di difoni, che utilizza la "voce" dello stesso parlatore.

Il modello a mistura di gaussiane, addestrato sui coefficienti *mel-cepstrali* estratti dal segnale neutro, è utilizzato per dividere questo spazio acustico in classi fonetiche equivalenti e per calcolare, per ogni classe identificata, i parametri delle funzioni di conversione.

Il metodo di trasformazione del segnale nel dominio delle frequenze ha fornito delle ottime prestazioni, come è stato dimostrato da un test percettivo in cui un segnale neutro convertito è stato riconosciuto come "arrabbiato".