

ESTRAZIONE DELLA FREQUENZA FONDAMENTALE IN UN CONTESTO DI MULTIMICROFONIA DISTRIBUITA

Federico Flego, Maurizio Omologo

ITC-irst, 38050 Povo (TN), Italy

RIASSUNTO

Questo lavoro parte dalla considerazione che il problema del calcolo della frequenza fondamentale F_0 è ancora in buona parte irrisolto nei casi in cui si debbano elaborare segnali vocali rumorosi e riverberanti.

Possibili scenari futuri, nei quali è necessario stimare con robustezza la frequenza fondamentale del parlatore, riguardano l' "ambient intelligence" basata su un esteso impiego di sensori (telecamere, microfoni, etc.). In questo ambito, la rimozione di vincoli nel posizionamento dei microfoni nell'ambiente, pur implicando una maggiore flessibilità per applicazioni come lo "speaker tracking" od il riconoscimento vocale a distanza (ovvero in condizioni di parlatore lontano alcuni metri dai microfoni), introduce nuove problematiche come ad esempio l'impossibilità di impiegare tecniche tradizionali di elaborazione basate sull'impiego di schiere di microfoni (es. beamforming). Per applicare tali tecniche, si dovrebbero infatti soddisfare condizioni specifiche nella geometria della schiera, in particolare nella distanza tra i microfoni adiacenti.

Nel presente lavoro, viene considerato uno scenario nel quale il parlatore si trova in un ambiente avente al suo interno una rete di microfoni distribuita arbitrariamente (Distributed Microphone Network).

L'elaborazione dei segnali multi-canale acquisiti tramite la suddetta rete di microfoni ha come scopo generale quello di derivare informazioni necessarie per la localizzazione del parlatore, per la classificazione di eventi acustici e per il riconoscimento finale della frase pronunciata. In un simile contesto, un possibile approccio all'estrazione della frequenza fondamentale puo' basarsi sull'analisi indipendente di ogni canale e la successiva applicazione di un metodo tipo "majority vote" che fornisca un'ipotesi di F_0 , frame per frame. Alternativamente, si puo' sfruttare la ridondanza di informazione fornita dai diversi canali, elaborando congiuntamente i diversi segnali per ottenere una stima piu' robusta di F_0 .

Questo contributo approfondisce l'utilizzo della tecnica di stima di F_0 basata sulla funzione Weighted Autocorrelation (WAUTOOC) estesa al contesto multi-canale.

Alla base di questo approccio sta la considerazione che il segnale acquisito da ogni microfono costituisce un'osservazione diversa di una stessa realizzazione vocale. Cio' permette di adottare un modello comune per la sorgente sonora e di elaborare congiuntamente i diversi flussi d'informazione.

Gli esperimenti condotti su una base di dati reale hanno dimostrato l'effettiva validità del metodo proposto per l'estrazione di F_0 da dati audio multi-canale. La sua efficacia viene dimostrata da diversi test sperimentali.

SESSIONE

In ordine di preferenza:

8) Elaborazione del Segnale VOCE

4) Prosodia (Durata, Intonazione)

30) Nuove Applicazioni e Sistemi per il TAL