

MISURA OGGETTIVA DELL' INTELLIGIBILITÀ IN APPLICAZIONI FORENSI BASATA SULLO STI

Giovanni Costantini^{1,2}, Massimiliano Todisco¹, Andrea Paoloni³

¹ Dipartimento di Ingegneria Elettronica, Università di Roma "Tor Vergata"
via del Politecnico, 1 - 00133 Roma

² Istituto di Acustica e Sensoristica "O.M. Corbino" - IDASC del CNR
Via del Fosso del Cavaliere, 100 - 00133 Roma,

³ Fondazione "Ugo Bordoni"
viale del Policlinico, 147 - 00161 Roma

Le intercettazioni ambientali sono diventate la fonte più frequente di elementi di prova nel processo penale, ma per utilizzarle è indispensabile la trascrizione nel testo scritto. Purtroppo, quando il segnale è poco comprensibile, potrebbe accadere che il verbale non corrisponde a qualcosa che è stato detto e le trascrizioni, in molti casi, riflettono più le opinioni del trascrittore che quelle di chi ha parlato. In molte occasioni il discorso è quasi incomprensibile, tuttavia, in alcune parti è possibile trarre dal segnale una corretta interpretazione. Le difficoltà di ottenere una buona trascrizione in testo scritto sono dovute principalmente ad un segnale inintelligibile a causa del basso livello di voce e del conseguente prevalere del rumore ambientale, spesso generato proprio allo scopo di mascherare la conversazione. Inoltre, come è ben noto ai linguisti, è quasi impossibile convertire il parlato in testo scritto senza perdere informazioni [1].

Nelle applicazioni forensi è aperto il dibattito su quale sia il limite nella qualità del segnale disponibile per dichiararlo non trascrivibile. In questo quadro riveste particolare importanza la possibilità di avere una misura oggettiva dell'intelligibilità per ricavarne stime sull'utilizzabilità di alcune registrazioni ai fini processuali.

Il segnale vocale può essere costituito da frasi, parole o semplici suoni senza significato (logatomi). Molti metodi e misure oggettive di intelligibilità del parlato sono stati proposti in passato [2, 3]. La maggior parte della letteratura in questo settore proviene dal mondo IT, dove il problema è quello di studiare l'influenza del canale di trasmissione sul segnale vocale [4, 5]. Gli algoritmi proposti spesso usano un approccio di tipo Double-Sided, ovvero operano sulla base di un confronto tra il segnale vocale pulito e il segnale trasmesso. Questo approccio non è utilizzabile in un contesto forense, perché il perito ha solo la versione rumorosa del segnale, quindi è necessario valutare l'intelligibilità con un approccio di tipo Single-Sided, cioè sulla base del solo segnale rumoroso.

Nel presente lavoro si è sviluppato un sistema basato sullo Speech Transmission Index (STI) [6] per la misura dell'intelligibilità del segnale vocale di tipo "Single-Sided". Per poter validare il sistema proposto è stato necessario disporre di un corpus di cui sia nota l'intelligibilità. In particolare sono state utilizzate 50 parole italiane rimate estratte da SAM EUROM 1 [7], con o senza significato, precedute dalla parola "PRENDI" e seguite dalla parola "INTANTO" lette da quattro voci diverse, due uomini e due donne. Questo corpus è stato poi reso rumoroso utilizzando rumore additivo. In particolare, al corpus è stato aggiunto rumore di tipo Pink, Hammer e Babble. Ogni tipo di rumore è stato aggiunto in cinque diversi gradi di rapporto segnale / rumore in dB ($S/N = 2, 0, -2, -4, -6$). Al termine delle operazioni, pertanto, si ottengono 60 diversi corpora (4 parlanti, 3 tipologie di rumore, 5 livelli di rumore) ognuno formato da 50 parole diverse.

Secondo la teoria dello STI, l'intelligibilità del parlato è correlata con la conservazione delle differenze spettrali tra fonemi in successione. Questo fenomeno può essere messo in relazione con la funzione involuppo che è determinata dalla sequenza specifica dei fonemi della frase. La misura basata sullo STI [8] è descritta schematicamente nella Fig. 1.

Nella Fig. 2 sono mostrati i risultati delle prove soggettive e delle misure oggettive del nostro sistema in funzione dell'SNR sui tre diversi rumori.

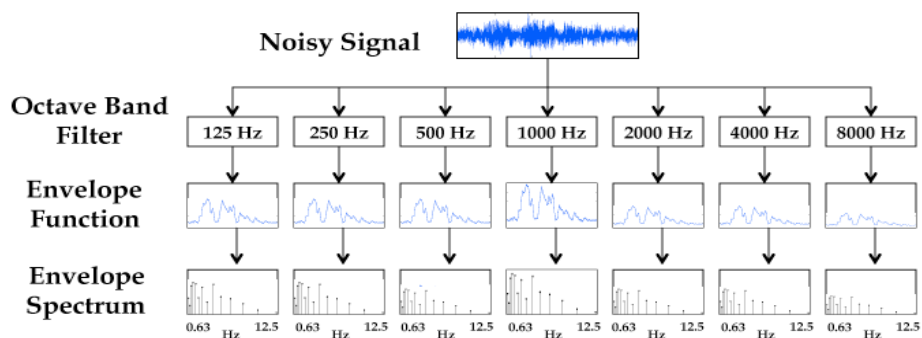


FIGURA 1: Diagramma a blocchi delle misure basate sullo STI

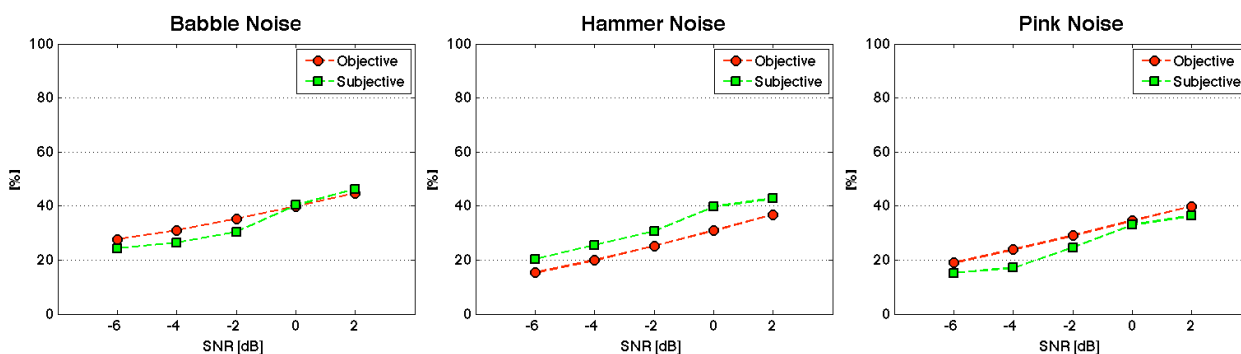


FIGURA 2: Misure soggettive (verde) e oggettive (rosso) a confronto

Riferimenti

- [1] H.Freser, "Issues in transcription: factors affecting the reability of transcripts as evidence in legal cases", *Fotensic linguistics*, vol. 102 (2002).
- [2] Ma J., Hu y., Loizou C.: "Objective measures for predicting speech intelligibility in most conditions based on new band importance functions" *JASA* 125, May 2009.
- [3] Nobuhiko Kitawaki, and Takeshi Yamada "Subjective and Objective Quality Assessment for Noise Reduced Speech", *ETSI Workshop on Speech and Noise in Wideband Communication*, May 2007, Sophia Antipolis, France
- [4] W. M. Liu, K. A. Jellyman, N. W. D. Evans, and J. S. D. Mason "Assessment of Objective Quality Measures for Speech Intelligibility", *INTERSPEECH 2008*, 9th Annual Conference of the International Speech Communication Association Brisbane, Australia September 22-26, 2008
- [5] Bianchi F., Paoloni A., Usai P., Vincenti A., "Misure di qualità di trasmissione effettuate su collegamenti telefonici terrestri via satellite Sirio" *FUB Rel.* 17, 1981
- [6] Payton K. L. "A method to determine the speech transmission index from speech waveforms", *JASA* 106, 3637-3648, 1999.
- [7] Chen D., Fourcin A., et alii, "EUROM A spoken language resource for the EU", *ESCA EUROSPEECH '95* Madrid September 1995.
- [8] Giovanni Costantini, Andrea Paoloni, Massimiliano Todisco, *Objective Speech Intelligibility Measures Based on Speech Transmission Index for Forensic Applications*, 39th International AES Conference on Audio Forensics: Practices and Challenges, Hillerød, Denmark, June 17–19, 2010, pp. 182-188.