



Associazione Italiana di Acustica  
XXVII Convegno Nazionale  
Genova, 26-28 maggio 1999

## ESTRAZIONE DI $f_0$ E RIDUZIONE DEL RUMORE MEDIANTE TECNICHE DI ANALISI BASATE SU MODELLI UDITIVI

P. Cosi<sup>(1)</sup>, Stefano Pasquin e Enrico Zovato<sup>(2)</sup>

<sup>(1)</sup> Istituto di Fonetica e Dialettologia – C.N.R.

Via G. Anghinoni, 10 – 35121 Padova (ITALY)

e-mail: [cosi@csrf.pd.cnr.it](mailto:cosi@csrf.pd.cnr.it) – www: <http://www.csrf.pd.cnr.it>

<sup>(2)</sup> Università di Padova, Dipartimento di Elettronica e Informatica

Via G. Gradenigo, 6 - 35100 Padova (ITALY)

www: <http://www.dei.unipd.it>

### SOMMARIO

Viene descritto un nuovo metodo per l'estrazione della frequenza fondamentale ( $f_0$ ) e per la *ripulitura* di segnale vocale rumoroso basato sull'applicazione del modello cocleare di Lyon, ed in particolare, sulla particolare elaborazione denominata *correlogramma*. Il correlogramma consente di calcolare il valore del ritardo di autocorrelazione per il quale i segnali in uscita dai canali cocleari del modello di Lyon hanno la stessa periodicità e consente quindi di calcolare l'andamento di  $f_0$ .

Nel caso di un rumore stazionario sovrapposto al segnale è possibile, per migliorare l'affidabilità del calcolo di  $f_0$ , applicare, nel dominio del correlogramma, una tecnica simile a quella denominata "*sottrazione spettrale*", che per tale analogia prende il nome di "*sottrazione nel correlogramma*". La stima di  $f_0$  viene, infatti, effettuata solo dopo aver sottratto il correlogramma medio del solo rumore al correlogramma del segnale rumoroso. L'andamento di  $f_0$ , anche nel caso di segnali rumorosi caratterizzati da rapporti segnale/rumore di 0 dB, risulta assai preciso ed affidabile.

La stessa tecnica di sottrazione nel dominio del correlogramma è utilizzata inoltre, previa introduzione di alcune modifiche sulla struttura del banco di filtri del modello di Lyon dettate da specifiche considerazioni energetiche, per ripulire e ri-sintetizzare, mediante una tecnica d'inversione, segnali vocali corrotti da rumore stazionario.

### INTRODUZIONE

La *frequenza fondamentale* ( $f_0$ , *pitch*)<sup>1</sup> è un parametro che racchiude in se gran parte dell'informazione prosodica veicolata dal segnale verbale e come tale gioca un

---

<sup>1</sup> Nel corso di questo lavoro, "frequenza fondamentale" o " $f_0$ " e "*pitch*" sono considerati sinonimi nonostante il loro diverso significato fisico e psicofisico.

ruolo molto importante nel processo di comunicazione. La corretta determinazione del *pitch* è quindi di fondamentale importanza in molte applicazioni sia nel campo dell'analisi, sintesi e riconoscimento del segnale verbale, che in quello della didattica per l'insegnamento delle lingue, come anche in quello relativo alle tecniche per la rieducazione del linguaggio.

Il sistema uditivo è intrinsecamente non lineare, ed è quindi difficile caratterizzare un modello in termini di variazioni d'ampiezza e di fase. L'informazione sulla periodicità di un segnale sembra essere però una caratteristica stabile e robusta che rimane pressoché indisturbata, anche attraverso le *non-linearità* imposte dal sistema.

Vi sono innumerevoli metodi in letteratura per l'estrazione di  $f_0$  [1], ma un algoritmo assolutamente affidabile rimane ancora difficile da ottenere soprattutto in condizioni particolarmente rumorose in cui il rapporto segnale/rumore si avvicina a 0 dB. Fra i vari metodi proposti, uno dei più vecchi e universalmente conosciuti è quello che si basa sull'analisi dell'*autocorrelazione a breve termine* (*STA*, *Short Time Autocorrelation* [2]) in cui  $f_0$  è calcolato individuando il ritardo temporale corrispondente al picco più elevato della funzione d'autocorrelazione a breve termine di un'opportuna finestra temporale del segnale. Il metodo proposto in questo lavoro si basa anch'esso su *STA*, ma questa è applicata non al segnale verbale, ma al *cocleagramma*, cioè alle uscite di tutti i canali del modello uditivo noto in letteratura come modello cocleare di Lyon [3-4]. La risultante rappresentazione tridimensionale, chiamata *correlogramma* [5], consente di visualizzare il segnale vocale in termini di tempo, frequenza e ritardo d'autocorrelazione. Il correlogramma consente di visualizzare le zone di frequenza maggiormente significative in termini di contenuto energetico e consente, inoltre, di calcolare il valore del ritardo d'autocorrelazione per il quale i segnali in uscita dai canali cocleari del modello di Lyon hanno la stessa periodicità. In altre parole, la più diretta informazione che si può trarre mediante il correlogramma, è quella relativa alle variazioni temporali del *pitch* del segnale vocale in ingresso.

Licklider propose già nel 1951 [5] il correlogramma come modello di percezione del *pitch* (*duplex theory of pitch perception*), ma questa tecnica per lungo tempo non fu presa seriamente in considerazione come procedimento computazionale per la rilevazione di  $f_0$  a causa della sua complessità. Solo nel 1984 fu pubblicata da Lyon [6] la prima immagine del correlogramma e alcuni ricercatori incominciarono ad utilizzarlo efficacemente per l'estrazione di  $f_0$  [7]. L'elevato costo computazionale è tuttavia compensato dall'affidabilità ottenibile nella stima di  $f_0$  anche in condizioni sfavorevoli, come in presenza di rumore di elevata intensità (0 dB SNR).

L'algoritmo proposto in questo lavoro si basa interamente sul correlogramma il quale viene però calcolato solo dopo aver applicato al segnale in ingresso un'opportuna tecnica di ripulitura del segnale per aumentare l'affidabilità dell'estrazione nei casi di basso rapporto segnale/rumore. Il correlogramma medio del solo rumore viene, infatti, sottratto al correlogramma del segnale rumoroso e solo successivamente ne viene calcolata la frequenza fondamentale. Per analogia con la più conosciuta tecnica denominata "*sottrazione spettrale a breve termine*" (*STSS*, *Short Term Spectral subtraction*) [8], questa tecnica di ripulitura nel dominio del correlogramma è stata denominata in [9] "*sottrazione sull'correlogramma a breve termine*" (*STCS*, *Short Term Correlogram Subtraction*).

La stessa tecnica di ripulitura è stata successivamente applicata con successo, assieme al proposto algoritmo di estrazione di  $f_0$ , ad una nuova tecnica di ri-sintesi

pilotata da un strategia di inversione [10-11] per ripulire un segnale fortemente degradato da un rumore stazionario di elevata potenza.

### IL FRONT-END Uditivo: "COCLEAGRAMMA" E "CORRELOGRAMMA"

Come indicato nell'introduzione, l'elaborazione del segnale in ingresso è effettuata mediante il modello cocleare di Lyon il cui risultato, detto cocleagramma, è una rappresentazione bidimensionale che simula le elaborazioni effettuate dalla coclea sul suono. Un banco di filtri passa-basso modella la propagazione dell'onda di pressione sonora lungo la **membrana basilare** (*BM*, *basilar membrane*) e la conversione dell'onda di pressione nella velocità della *BM* (**filtraggio**). Un meccanismo non lineare di rilevazione, realizzato mediante un *raddrizzatore a semionda* o (*HWR*, *Half Wave Rectifier*), converte la velocità della *BM* in una rappresentazione del numero medio di impulsi nervosi (**rivelazione**) indotti nelle cellule cigliate dal movimento della *BM*, e un controllo automatico di guadagno (*AGC*, *Automatic Gain Control*) adatta il guadagno del sistema in risposta al livello di attività, in modo da comprimere l'ampia gamma di ampiezze del suono in ingresso in un intervallo limitato di intensità delle scariche neurali. Lo schema completo del modello di Lyon è illustrato in Figura 1 assieme al cocleagramma corrispondente ad una vocale /a/ pronunciata da un parlante maschile. Gli impulsi neurali prodotti dai ricettori presenti in prossimità della coclea vengono trasmessi per mezzo del nervo acustico alla corteccia cerebrale e qui subiscono ulteriori elaborazioni. Un'ipotesi "accettabile" per una simulazione delle operazioni che sono eseguite a livello cerebrale è rappresentata dal calcolo di una serie di autocorrelazioni sulle sequenze di impulsi provenienti dalle cellule cigliate afferenti a diverse porzioni della coclea. In tal modo viene conservata sia l'informazione sulla frequenza, sia quella temporale. Il correlogramma è una particolare elaborazione eseguita sugli *N* canali uscenti dal cocleagramma che rende conto di ciò che potrebbe avvenire a livello corticale e si ottiene calcolando le autocorrelazioni nel dominio del tempo degli *N* segnali che formano il cocleagramma. Poiché l'autocorrelazione è essa stessa funzione di una terza variabile, il correlogramma è dunque una rappresentazione tridimensionale del segnale vocale, come funzione del tempo, della frequenza (cioè della posizione cocleare), e del ritardo di autocorrelazione. In tal modo se un segnale presenta delle periodicità, queste vengono messe in evidenza dalla autocorrelazione del segnale stesso; infatti, essa presenta un massimo locale ogniqualvolta il ritardo di autocorrelazione (*time lag*)  $\tau$  è pari a un multiplo intero del periodo del segnale. I segnali di interesse per la nostra analisi, come il parlato, non sono periodici, e più in generale non sono nemmeno stazionari, per questo l'operazione di autocorrelazione è eseguita su finestre temporali di durata fissa e traslate di un intervallo di tempo costante (*STA*). Entro queste finestre il segnale, oltre ad essere considerato stazionario, può talvolta esibire delle periodicità che sono facilmente individuabili nel correlogramma. In questo lavoro è stata utilizzata una finestra di Hamming modificata, che offre i vantaggi di riduzione degli effetti ai bordi e, dualmente, di una maggiore risoluzione in frequenza, ed in più, offre altri vantaggi computazionalmente utili per le procedure di ri-sintesi [12].

Se la finestra di analisi  $w(n)$  ha lunghezza  $L$ , l'autocorrelazione  $r_x$  a breve termine di  $x_i(n)$ , uscita dell'*i*-esimo canale del cocleagramma, è rappresentata da (1), mentre con  $C_i$  in (2) si indica il correlogramma campionato nel tempo ogni  $S$  campioni. Nelle nostre elaborazioni il segnale è campionato a 16 kHz e il correlogramma è calcolato su finestre temporali di lunghezza  $L=512$  campioni, pari a  $t=32$ ms e i frame di correlogramma sono

calcolati ogni  $S=128$  campioni, pari a  $t=8$ ms. Essendo l'autocorrelazione di un segnale una sequenza pari, in ogni frame tale funzione è illustrata solo per ritardi positivi.

$$r_{x_i}(m, \tau) = \frac{1}{L} \sum_{n=0}^{L-|\tau|-1} x_{w_i}(n, m) x_{w_i}(n - |\tau|, m) \quad i=1, \dots, N \quad (1)$$

$$c_i(mS, \tau) = r_{x_i}(mS, \tau) \quad i=1, \dots, N \quad (2)$$

$$x_{w_i}(n, m) = x_i(n) \cdot w(m - n) \quad \tau : \text{ritardo di autocorrelazione} \quad (3)$$

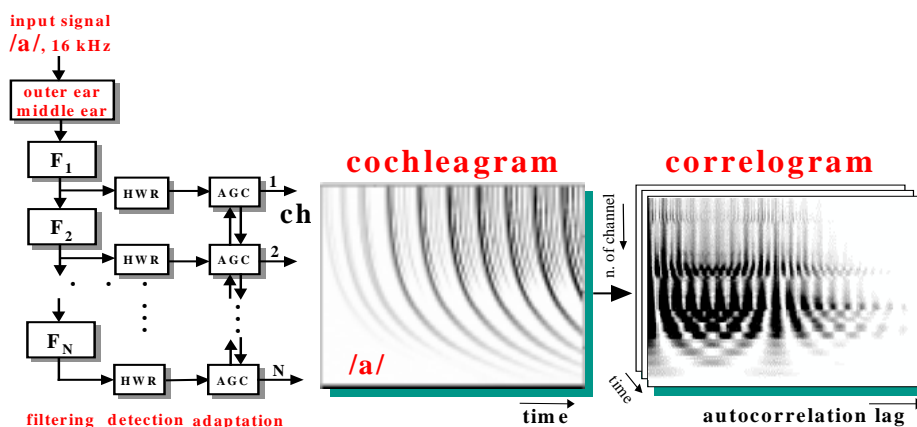


Figura 1. Diagramma a blocchi del modello cocleare di Lyon, del cocleogramma e del correlogramma relativi alla vocale /a/ pronunciata da un parlante maschile.

### ESTRAZIONE DEL PITCH

Il correlogramma, come mostrato in Figura 1, non è una rappresentazione statica ma dinamica. Poiché l'informazione sulla periodicità di un segnale rimane pressoché invariata lungo tutte le trasformazioni che avvengono nei canali cocleari del modello uditivo di Lyon, se un suono è periodico, le funzioni di autocorrelazione di un elevato numero di canali indicheranno tutte un picco più elevato in corrispondenza del suo periodo fondamentale  $\tau_0 = 1/f_0$ . Come introdotto in [6] viene eseguita una somma sui valori del correlogramma  $cor_i(\tau)$  lungo tutti i canali per ottenere la funzione  $SC(\tau)$  indicata in (4) denominata *somma sul correlogramma (SC, Summary Correlogram)*, dove il picco in corrispondenza di  $\tau_0$  risulta enfatizzato, come illustrato in Figura 2.

$$SC(\tau) = \sum_{i=1}^N cor_i(\tau) \quad i=1, \dots, N \quad (4)$$

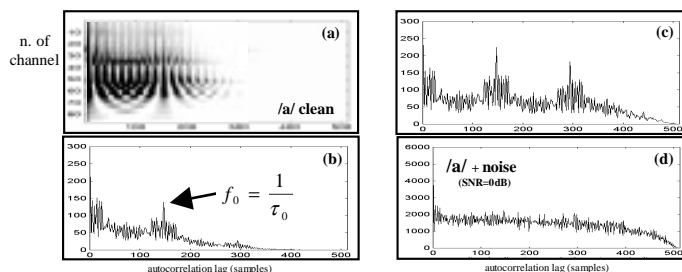


Figura 2. *Somma nel dominio del correlogramma (SC, Summary Correlogram)* relativo al correlogramma (a) della vocale /a/ di un parlante maschile senza (b) e con (c) normalizzazione. In (d) è illustrato il SC nel caso della stessa vocale con rumore sovrapposto a 0 dB SNR.

Poiché nel calcolo delle *STAs* il segnale in analisi viene 'finestrato', come indicato in (3), prima di calcolare  $SC(\tau)$  si rende necessaria la normalizzazione di  $cor_i(\tau)$  espressa in (5) al fine di uniformare il livello di significatività dei picchi delle *STAs* che altrimenti seguirebbero un andamento simile all'autocorrelazione della finestra.

$$cor'_i(\tau) = \frac{r_{x_i}(\tau)}{r_w(\tau)} \quad i=1, \dots, N \quad (5)$$

Quando il segnale in ingresso è fortemente corrotto da rumore la procedura è stata modificata in quanto il *SC* spesso non esibisce picchi significativi come esemplificato in (d) nella Figura 2. Nel caso di un rumore stazionario sovrapposto al segnale vocale, prima di calcolare  $f0$ , è applicata nel dominio del correlogramma una tecnica simile a quella denominata "sottrazione spettrale a breve termine" (*STSS*) [8], che per tale analogia prende il nome di "sottrazione sul correlogramma a breve termine" (*STCS*) [9]. Prima di calcolare  $f0$ , infatti, il correlogramma medio del solo rumore calcolato automaticamente negli istanti di silenzio mediante la formula espressa in (6) viene sottratto al correlogramma del segnale rumoroso.

$$rif_i^{noise}(\tau) = \frac{1}{N} \sum_{m=1}^N cor_i^{noise}(m, \tau) \quad i=1, \dots, N \quad (6)$$

Nonostante il fatto che le trasformazioni non lineari implicite nel front-end uditivo non giustifichino l'applicazione della "sovrapposizione degli effetti", la procedura di sottrazione nel dominio del correlogramma si è rivelata assai efficace. Per dimostrare l'applicabilità della sovrapposizione degli effetti e quindi della sottrazione spettrale nel dominio del correlogramma sono state definite le grandezze  $S_i$  espresse in (7) denominate 'somme sul ritardo di autocorrelazione' (*STL, Sum on the Time Lag*) e si è verificato sperimentalmente [10], sia che queste grandezze rappresentano il "grado di eccitazione" (*ED, Excitation Degree*) del segnale su ogni singolo canale, nel frame considerato, cioè sono in un rapporto di proporzionalità con le vere e proprie 'energie' nei vari canali, sia che nel loro dominio l'applicazione della sovrapposizione degli effetti è largamente giustificata.

$$S_i = \sum_{\tau} cor_i(\tau) \quad i=1, \dots, N \quad (7)$$

Si vede infatti che  $SN_i \cong S_i + N_i$ , dove le tre *STLs* si riferiscono rispettivamente al grado di eccitazione dei canali corrispondenti al segnale 'rumoroso' ( $SN_i$ ), al segnale 'pulito' ( $S_i$ ) ed al rumore sovrapposto ( $N_i$ ). Come illustrato in Figura 3, nel dominio delle *STLs* l'applicazione della sovrapposizione degli effetti per ricostruire l'energia del segnale pulito  $\hat{S}_i \cong SN_i - N_i$  è sperimentalmente accettabile.

La descrizione statistica del rumore calcolata in (6) risulta dalla media aritmetica del correlogramma del rumore sovrapposto calcolato sui singoli frame, ed è solo una prima approssimazione del reale correlogramma del rumore in ogni frame. Di conseguenza le *STLs* calcolate sul rumore mediante (7)  $\hat{N}_i = \sum_{\tau} rif_i^{noise}(\tau) \approx N_i$  sono solo una stima delle 'energie' del rumore nei vari canali frame per frame ( $N_i$ ), da cui risulta che una nuova versione della *somma sul correlogramma* denominata *somma robusta sul correlogramma* (*RSC, Robust Summary Correlogram*), per il calcolo di  $f0$  nel caso di segnale fortemente corrotti da rumore, può essere espressa mediante (8a,8b,8c).

$$\hat{S}_i \cong SN_i - \hat{N}_i \quad \hat{N}_i = \sum_{\tau} rif_i^{noise}(\tau) \approx N_i \quad (8a)$$

$$\hat{S}_i^d = \max(0, SN_i - d\hat{N}_i) + \varepsilon \quad d=[1,2] \quad \text{or} \quad \hat{S}_i^d = \max(0, SN_i - (\hat{N}_i + 3\sigma_i)) + \varepsilon \quad (8b)$$

$$SC_w(\tau) = \sum_{i=1}^N \hat{S}_i^d \frac{cor_i(\tau)}{\sum_{\tau} cor_i(\tau)} \quad RSC(\tau) = \frac{SC_w(\tau)}{r_w(\tau)} \quad i=1, \dots, N \quad (8c)$$

Le nuove grandezze  $\hat{S}_i^d$  in 8b forzano le energie stimate ad essere sempre maggiori di zero ovviando al problema posto dalle possibili fluttuazioni di  $N_i$  rispetto a  $\hat{N}_i$  e dal fatto che l'applicazione della sovrapposizione degli effetti alle energie dei segnali che passano attraverso varie trasformazioni lineari è valida solo in prima approssimazione, condizioni che potrebbero far verificare  $\hat{S}_i \equiv SN_i - \hat{N}_i \leq 0$ . Il caso  $d=2$  in (8b) garantisce una migliore ripulitura ma ha lo svantaggio di eliminare anche alcune parti utili dei canali dove  $S(i)$  è significativa. Nel caso di segnali fortemente corrotti da rumore le grandezze  $\hat{S}_i^d$  consentono la discriminazione dei canali più corrotti da quelli meno corrotti, vale a dire, in definitiva, con la *somma robusta sul correlogramma* (RSC) i canali meno corrotti dal rumore contribuiscono maggiormente al calcolo finale di  $f_0$ .

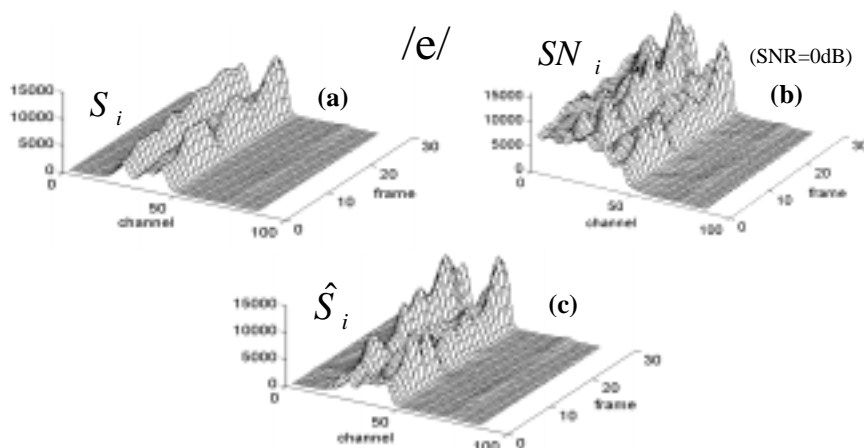


Figura 3. Visualizzazione delle 'somme sul ritardo di autocorrelazione' (STLs), relative ad una vocale /e/ pronunciata da un parlante maschile, nel caso di segnale 'pulito' (a), 'rumoroso' (0 dB) (b) e 'ricostruito' dopo la sottrazione con  $\hat{S}_i \equiv SN_i - N_i$ .

In Figura 4 è illustrato il diagramma a blocchi completo del nuovo sistema per l'estrazione del *pitch* denominato **CS-RPE** (*Correlogram Subtraction – Robust Pitch Extraction*) basato sul RSC. Prima del calcolo finale di  $f_0$ , per aumentare la correttezza dei valori stimati, è applicato un algoritmo di correzione degli errori basato sulle zone di più elevata affidabilità ("island driven") per quanto riguarda i valori di  $f_0$ . Dapprima si individuano  $I_k$  zone stabili di  $f_0$ , dove con stabili si intendono quelle zone con almeno D (attualmente  $D=8$ ) valori consecutivi di *pitch* le cui differenze sono inferiori ad una determinata soglia  $\Delta f$  (attualmente  $\Delta f = |30\text{Hz}|$ ), successivamente si correggono gli errori isolati (eventuali errori di raddoppio, dimezzamento o errori più grossolani) all'interno di queste regioni stabili, mediante filtri di *smoothing* a media mobile e mediante un'opportuna euristica, e, da ultimo, si uniscono insieme queste regioni mediante una procedura di continuità che ricerca i valori possibili di  $f_0$  frame per frame utilizzando il RSC nel tentativo di individuare possibili picchi target.

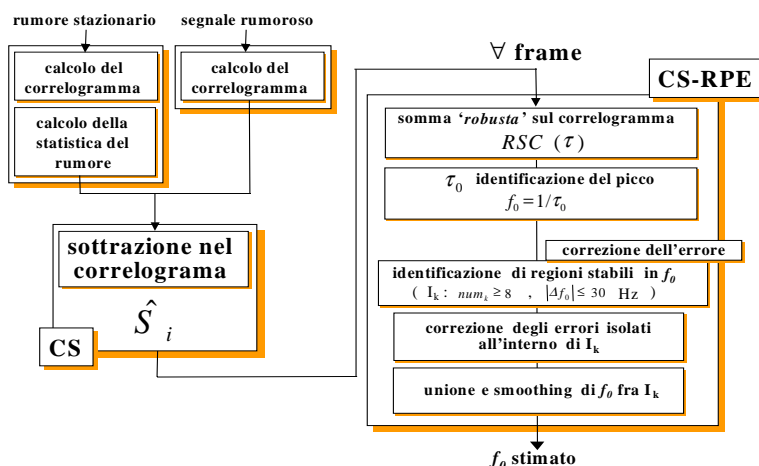


Figura 4. Diagramma a blocchi dell'algoritmo 'robusto' per l'estrazione del *pitch* (CS-RPE, *Correlogram Subtraction – Robust Pitch Extraction*).

Un esempio dell'applicazione dell'algoritmo CS-RPE ad un segnale fortemente corrotto da rumore stazionario (0dB SNR) è illustrato in Figura 5 (5a,b,c), dove l'efficacia dell'estrazione è chiaramente riconoscibile.

### RIPULITURA DEL SEGNALE RUMOROSO

La stessa tecnica può essere applicata per ripulire e risintetizzare un segnale pulito a partire da un segnale fortemente corrotto da rumore. Prima di applicare le procedure di inversione descritte in [10-12] si può, infatti, applicare la STCS per ricostruire un correlogramma pulito *corSNC* a partire da un correlogramma di un segnale rumoroso *corSN*. Mediante la formula indicata in (9) è, infatti, possibile,  $\forall$  frame  $m$ , ricostruire un correlogramma dove i canali più corrotti dal rumore siano fortemente attenuati.

$$corSNC_i(\tau) = \hat{S}(i) \frac{corSN_i(\tau)}{\sum_{\tau} corSN_i(\tau)} \quad \hat{S}_i = \sum_{\tau} corSNC_i(\tau) \quad \forall \text{ frame } m \quad i=1, \dots, N \quad (9)$$

Il correlogramma ricostruito è poi forzato ad appartenere ad un vero e proprio segnale periodico di frequenza fondamentale  $f_0$ , precedentemente calcolata mediante l'algoritmo CS-RPE. Questo è ottenuto, operando nel dominio del correlogramma, mediante l'applicazione di un'opportuna finestra *anti-aliasing* sui singoli canali, nel dominio del ritardo di autocorrelazione, per isolare il periodo target  $\tau_0 = 1/f_0$  estratto dall'algoritmo CS-RPE e mediante l'applicazione di un operatore di ripetizione periodica [9]. Il correlogramma ottenuto è infine inviato all'algoritmo di risintesi mediante inversione [10-11] per la generazione del segnale ripulito. Un esempio dell'applicazione dell'algoritmo di ripulitura è illustrato in Figura 5 (5d,e,f).

### COCLUSIONI

In questa lavoro è stata introdotta la procedura denominata *sottrazione nel correlogramma*, mediante la quale è stata resa possibile una stima robusta del *pitch* di segnali vocali in ambiente fortemente rumoroso (SNR=0dB). La stessa procedura ha anche permesso di ottenere, per mezzo degli algoritmi di inversione del correlogramma, un segnale sintetizzato percettivamente più gradevole e più intelligibile del corrispondente rumoroso. Il risultato raggiunto compensa il limite computazionale

dettato dall'elevata complessità dell'algoritmo anche in luce di possibili applicazioni future quali ad esempio la separazione di due segnali vocali sovrapposti.

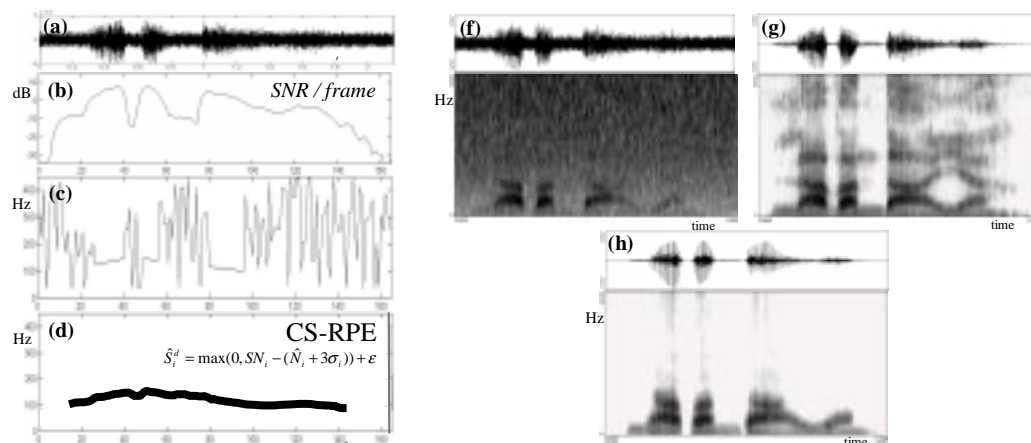


Figura 5. Relativamente allo stimolo rumoroso /lavan'daja/ (a), sono illustrati: in (b) il rapporto segnale rumore (SNR), frame per frame (SNR globale di 0dB); in (c) la stima ottenuta mediante l'algoritmo originale senza l'applicazione della tecnica di *sottrazione sul correlogramma (STCS)*, in (d) la stima calcolata con l'algoritmo proposto **CS-RPE**. In (h) sono illustrati la forma d'onda e lo spettrogramma del segnale ripulito, risultato dell'applicazione dell'algoritmo di ri-sintesi applicato al segnale rumoroso (f) ottenuto 'sporcando' il segnale pulito (g) mediante un rumore stazionario (SNR 0dB).

## BIBLIOGRAFIA

- [1] H. Hess, *Pitch Determination of Speech Signals*, Springer-Verlag, Berlin, New York,, 1983.
- [2] L.R. Rabiner, *On the use of autocorrelation analysis for pitch detection*, Proc. IEEE, Vol. 58, 1970, pp.707-712.
- [3] R. F. Lyon, *A Computational Model of Filtering, Detection, and Compression in the Cochlea*, Proc. IEEE-ICASSP, 1982, pp.1282-1285.
- [4] F. Lyon, *A Computational Model of Binaural Location and Separation*, Proc. IEEE-ICASSP, Boston 1983, pp. 1148-1151.
- [5] J.C.R. Licklider, *A duplex theory of pitch perception*, Experientia 7, 128-133. Also reprinted in *Pysiological Acoustic* (E.D. Shubert ed.), Dowden, Hutchinson&Ross Inc. Stroudsburd, PA, 1979.
- [6] R. F. Lyon, *Computational Models of Neural Auditory Processing*", Proc. IEEE-ICASSP, S.Diego, 1984,36.1.1-36.1.4.
- [7] Slaney M. and Lyon R.F., "A Perceptual Pitch Detector", Proc. IEEE-ICASSP, 1990, 357-360.
- [8] A. Acero, *Acoustical and Environmental Robustness in Automatic Speech Recognition*. Kluwer Academic Publishers, Boston, 1993.
- [9] P. Cosi, S. Pasquin and E. Zovato, *Auditory Modeling Techniques for Robust Pitch Extraction and Noise Reduction*, Proc. ICSLP-98, International Conference on Spoken Language Processing, Sydney, Australia, 30 Nov. – 4 Dec., 1998, VII, 2807-2810.
- [10] M. Slaney, D. Naar and R.F. Lyon, *Auditory Model Inversion for Sound Separation*, Proc. IEEE-ICASSP, Adelaide, 1994, II.77-80.
- [11] Piero Cosi and Enrico Zovato, "Lyon's Auditory Model Inversion: a Tool for Sound Separation and Speech Enhancement", Proc. ESCA Workshop on 'The Auditory Basis of Speech Perception', Keele University, Keele (UK), 15-19 July, 1996, 194-197.
- [12] E. Zovato, *Sintesi di suoni mediante inversione di un modello uditivo*, Dip. Elettronica e Informatica, Univ. di Padova, Tesi di Laurea, A.A. 1994-1995.