Robust hybrid pitch detector for pathologic voice analysis

B. Boyanov, S. Hadjitodorov, B. Teston, D. Doskov

June 16, 17 & 18
16, 17 et 18 Juin
1997

Centre Hospitalier Universitaire « La Timone »
Marseille France

Organized by / *Organisé par*

Laboratoire Parole et Langage - Université de Provence
ESA CNRS 6057

Laboratoire d'Audio-Phonologie clinique - Université de la Méditerranée
Assistance Publique de Marseille

With / *Avec*

European Speech Communication Association (ESCA)

European Laryngological Society (ELS)

Groupe de Recherche Européen sur le Larynx (GREL)

Société Française d'Acoustique (SFA)
(Groupe Francophone de la Communication Parlée)

Société Française de Phoniatrie (SFP)

ROBUST HYBRID PITCH DETECTOR FOR PATHOLOGICAL VOICE ANALYSIS

*Boyanov, B., S. Hadjitodorov, B. Teston *, D. Doskov***
Center on Biomedical Engineering, Acad. G. Bonchev St, Block 105, 1113 Sofia, Bulgaria
** Lab."Parole et Language",URA 261 CNRS, 29, av. R. Schuman, Aix en Provence, France*
*** Phoniatric Dep. Univ. Hospital "Queen Jovanna" Sofia, BULGARIA*

ABSTRACT

A hybrid pitch period (To) detector characterized by parallel analyses of the speech signal in temporal, spectral and cepstral domains and preprocessing for periodic/aperiodic(unvoiced) separation (PAS) is proposed. The preprocessing is realized by analyses in these three domains and PAS by multilayer perceptron neural network. To evaluation is carried out by analyses in: 1. Time domain 1.1 Preprocessing by adaptive central clipping (ACC). 1.2. Calculation of the short-term autocorrelation function ($R_{st}(\tau)$) 1.3. Calculation of To by analysis of the distances between the maximas in $R_{st}(\tau)$ - LADM . 2. Spectral domain: 2.1. Spectral ACC. 2.2. Calculation of $R_{st}(\tau)$ over the power and logarithmic spectra. 2.3. To detection by LADM. 3. Cepstral domain: 3.1. Liftering the cepstra to eliminate the formant structure. 3.2 Calculation of $R_{st}(\tau)$. 3.3. To detection by LADM. Final To evaluation by a logical analysis of the results from the three domains. **EXPERIMENTS** Two phonations of the vowel "a" of 40 speakers and 62 patients were analyzed. For the proposed detector errors were significantly minimized.

RESUME

On propose un détécteur hybrid du pitch période (To), qui se caracterise par une analyse parallèle de la parole effectué dans les domaines temporel, spectral et cepstral et avec un traitement pour séparation des segments périodiques et apériodiques (PAS). Le traitement est realisé par analyse dans les trois domaines and PAS à l'aide d'un reseau multicouche neuromimétique. Le To est calculé par des analyses suivantes: 1. Temporel domaine: 1.1 Traitement avec clipping adaptive central (ACC). 1.2. Calcul de la fonction d'autocorrélation ($R_{st}(\tau)$) 1.3. Calcul du To par analyse des distances entre les maxima de $R_{st}(\tau)$ (LADM) . 2. Spectral domaine: 2.1. Spectral ACC. 2.2. Calcul de la $R_{st}(\tau)$ sur les spectres linéaires et logarithmiques. 2.3. Détection du To à l'aide de LADM. 3. Cepstral domaine: 3.1. Liftering du cepstrum pour éliminer la structure formantique. 3.2 Calculation de la $R_{st}(\tau)$ 3.3. Détection du To à l'aide de LADM. Calculation finale du To par une analyse logique des values du To des trois domains. **EXPERIMENTS** Deux phonations de la voyelle "a" de 40 locuteurs et 62 patients etaient analysées. Le nombre des erreurs commis par le détecteur propose etait minimisé.

INTRODUCTION

The basic problem in acoustical analysis of pathological voices is the correct To determination, because many voice parameters are calculated on the basis of To [2-5]. Many methods for To evaluation are developed [1,2], however the experimental researches [3,4,5] show that they detect To from pathological voices. with a significant number of errors. In order to minimize these errors a To detector, characterized by parallel analyses in temporal, spectral and cepstral domains and preprocessing for robust PAS is proposed.

PREPROCESSING -PAS

Analysis in time domain

The $R_{st}(\tau)$ is evaluated. However formants and noisy components causes difficulties in To evaluation by means of $R_{st}(\tau)$. To increase the accuracy the following algorithm for central clipping with dynamic threshold's adaptation (CCDAT) is proposed and used:

Step1. Center clipping (CC) of the signal, using thresholds for the maxima (TR_{max}) and minimas (TR_{min}) by the procedure [2].

Step 2. Calculation of the number ($Nmax_{ok}$) of positive samples and the number ($Nmin_{ok}$) of negative samples that are encoded by means of CC - i.e. samples fulfilling the amplitude selection with respect to TR_{max} and TR_{min}.

Step 3. Calculation of the ratio N_{ok}-to-N_{tot} for maxima (ROT_{max}) and minma (ROT_{min})

$$ROT_{max} = Nmax_{ok} / N_{tot} \quad (1a)$$

$$ROT_{min} = Nmin_{ok} / N_{tot} \quad (1b)$$

where: N_{tot} - number of all the samples.

Step 4. Verification the result from the CC:

Successful central clipping:

$$ROT1_{max} < ROT_{max} < ROT2_{max} \quad (2a)$$

$$ROT1_{min} < ROT_{min} < ROT2_{min}, \quad (2b)$$

where: $ROT1_{min} = ROT1_{max} = 0.2$;

$ROT2_{max} = ROT2_{min} = 0.6$; are coefficients determinated on the basis of the analysis of the speech of 20 patients.

Unsuccessful central clipping:

case 1. Very large thresholds (not enough encoded samples - loss of information):

a) over the maximas:

$$ROT_{max} < ROT1_{max} \quad (3)$$

In this case the value of TR_{max} is decreased:

$$TR_{max} = TR_{max} - p \quad (4)$$

where: $p = 0.05$ is experimentally determinated coefficient (20 patients analyzed).

b)the same procedure over the minimas

Go to step 1 - restart the CC with the new values of TR_{max} or/and TR_{min}.

case 2. Very low thresholds (too many encoded samples - as a result noises and formnats are not enough suppressed):

a) over the maximas:

$$ROT_{max} > ROT2_{max} \quad (5)$$

In this case the value of TR_{max} is increased:

$$TR_{max} = TR_{max} + p \quad (6)$$

where: $p = 0.05$.

b)the same procedure over the minimas

Go to step 1 - restart the CC with the new values of TR_{max} or/and TR_{min}.

Calculation of the $R_{st}(\tau)$

After successful CC the $R_{st}(\tau)$ is evaluated over the clipped signal. For PAS the following parameters are used: $R_{st}(\tau=0)$ and $R_{st}(\tau=To)$, corresponding to the largest peak in the range of To.

Analysis in spectral domain

The spectral autocorrelation function ($R_s(\tau)$) is calculated over the power ($R_{pow}(\tau)$) and logarithmic ($R_{log}(\tau)$) spectra. These spectra are preprocessed by means of the procedure: **Step 1.** Calculation of the group delay function (GDF) [4]. **Step 2.** Detection of the low (LER) and high energy regions (HER) by peak-picking. **Step 3.** CCDAT for every LER and HER. The $R_s(\tau)$ is calculated over the clipped power and logarithmic spectra. For the PAS are used: $R_{log}(\tau=0)$; $R_{log}(\tau=To)$; $R_{pow}(\tau=0)$ and $R_{pow}(\tau=To)$

Analysis in cepstral domain

The influence of the formant structure is minimized by means of liftering the cepstra. The $R_{st}(\tau)$ is calculation over this cepstra ($R_{cep}(\tau)$). For the PAS $R_{cep}(\tau=To)$ is used.

For the PAS the energies (Es) in spectral and in time domain (E) (bandwidth 90 - 3000 Hz) are used too.

PAS classification by means of MLP

The PAS is realized by MLP because [6] MLP are characterized by good discriminant capabilities and high classification power. A three layer MLP can form regions as complex as those formed using mixture distributions and nearest neighbor classifiers. For that reason such MLP(two hidden layers and one output layer) is used. The first layer consists of 20 neurons, the second of 2 neurons and the last of one output neuron. The MLP is trained by means of back-propagation algorithm. For periodic segments the output is one and for aperiodic it is zero.

PITCH PERIOD DETECTION

Analysis in time domain (detector TP)

The following algorithm for To evaluation by analysis of $R_{st}(\tau)$ is proposed and used:

Step 1. Calculation of a threshold:

$ATR = q\, R_{max}(\tau_{max})$ (7)

where: q - experimentally evaluated coefficient (here q = 0.6); $R_{max}(\tau_{max})$ - global maximum in $R_{st}(\tau)$ in the range of To: τ [2 - 15 ms]

Step 2. Search for maximas $R_m(\tau_s)$, s=1,.., S) larger than ATR and at longer distance than the shortest To (To_{short}) (here To_{short}=2 ms):

$R_m(\tau_s) = R_{st}(\tau)$ if $R_{st}(\tau) > ATR$ and

$|\tau_s - \tau_{s-1}| > To_{short}$ (8)

Step 3. If not one maximum is detected then τ_{max} is classified as the To

Step 4. If several $R_m(\tau_s)$ are detected then the time intervals (d_s) between them are calculated.

Step 5. Verification if these distances are nearly equal:

$|d_s - d_{s-1}| < 0.2\, d_s$ (9)

Step 6. Evaluation of To:

If inequality (9) is fulfilled for all $R_m(\tau_s)$ then:

$To = d_s$ (10)

else:

To_t =0 - To is not evaluated for the segment.

Analysis in spectral domain (detectors PS and LP)

In spectral domain two values of To (To_l and To_p) are evaluated by analysis of $R_s(\tau)$ calculated over the power (detector PS) and logarithmic (detector LS) spectra. The determination of To is realized by means of the algorithm used in time domain and if it fails then To_l =0 and/or To_p =0.

Analysis in cepstral domain (detector CEP)

In cepstral domain To (To_c) is evaluated by analysis of $R_{cep}(\tau)$. The determination of To is realized by means of the algorithm used in time domain and if it fails then To_c =0.

To evaluation for the segment

The To evaluation is realized on the basis of the different robustness of the detectors over pathological voices [4]:a) LS and PS are generally very reliable; b) TP is reliable for most of the pathological voices; c) CEP fails for many pathological voices. As result To is evaluated by means of the following logical analysis:

case 1: if $|To_t - To_p| < qTo_t$ and $|To_t - To_l| < qTo_t$ then $To = To_t$.

case 2: if $|To_t - To_p| < qTo_t$ and $|To_t - To_c| < qTo_t$ then $To = To_t$.

case 3: if $|To_t - To_l| < qTo_t$ and $|To_t - To_c| < qTo_t$ then $To = To_t$.

where: q = 0.3 - takes into account deviations in To caused by the discrete spectral structure

In all the other cases no decision about the value of To is made.

EXPERIMENTAL RESEARCH

Two phonations of the vowel "a" of 40 speakers and 62 patients were analyzed. The signals were quantified directly into the computer's memory at 20480 Hz and precision 12 bit using the "DSP Sonagraph 5500". To is evaluated by the following methods [2]: cepstral (CEP), clipped autocorrelation (CAC); SIFT using inverse filtering and spectral autocorrelation (SAF). The standard pitch contour is evaluated by visual inspection of the signal, the spectrum and the cepstrum. The following errors were observed: a) voiced segment classified as unvoiced (VU); b) unvoiced segment classified as voiced (UV); c) gross errors (G) deviation form the true To more than 10%;d) fine errors- deviation less than 10%. e) only for the proposed method - voiced segment where the method can not evaluate To (ND). :The percentage of the segments where errors are detected with respect to the total number of segments is shown on table 1. The results show that the proposed detector allows To determination from pathological voices with significantly minimized number of errors.

REFERENCES

[1] Boyanov B., S. Hadjitodorov, T. Ivanov, G. Chollet. (1993) Robust Hybrid Pitch Detector. Electronic Letters v.29, pp. 1924-1926.

[2] Hess, W. (1983),.Pitch determination of speech signals, Springer Verlag, Berlin,

[3] Laver J, Hiller S and Hanson R: (1982) Comparative performance of pitch detection

algorithms of dysphonic voices. In: *Proc. ICASSP'82*, vol. 1, pp. 123-127.

[4] Boyanov B. (1993), Analysis of pathological voice. Report under contract ERB-CIPA-CT-92-0170 with the European Community.

[5] Bielamowicz S., Kreiman J, Gerratt B, Dauer M, Berke G. (1996) Comparison of voice analysis systems for perturbation measurement. J.S.H.R., v. 39, pp. 126-143.

[6] Lippmann R. (1989), Pattern classification using neural networks, *IEEE Communications Magazine*, pp. 47-64

Table 1. Percentage of segments where errors are detected

Method	Type of voice	UV [%]	VU [%]	Gross [%]	Fine [%]	ND [%]
CAC	Norma	0.2	0.1	0.3	0.7	none
	Pathology	4.3	3.7	5.6	5.3	none
SAF	Norma	0.1	0.4	0.5	0.8	none
	Pathology	3.3	2.2	4.4	7.4	none
CEP	Norma	0.2	0.3	0.4	0.5	none
	Pathology	8.7	4.2	10.2	7.6	none
SIFT	Norma	0.3	0.9	1.1	0.6	none
	Pathology	7.3	6.7	8.3	7.1	none
proposed	Norma	0	0.5	0.4	0.7	0.3
	Pathology	0	0.8	0.7	4.1	2.8