

Atti  
-21-

## ATTI

1. *Il controllo terminologico delle risorse elettroniche in rete: tavola rotonda, Firenze 27 gennaio 2000*, a cura di Paola Capitani, 2001
2. *Commemorazione di Michele Della Corte*, a cura di Laura Della Corte, 2001
3. *Disturbi del comportamento alimentare: dagli stili di vita alla patologia*, a cura di Corrado D'Agostini, 2002
4. *Proceedings of the third International Workshop of the IFIP WG5.7 Special interest group on Advanced techniques in production planning & control : 24-25 February 2000, Florence, Italy*, edited by Marco Garetti, MarioTucci, 2002
5. *DC-2002: Metadata for E-Communities: Supporting Diversity And Convergence 2002: Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2002, October 13-17, 2002, Florence, Italy*, organized by Associazione Italiana Biblioteche [et al.], 2002
6. *Scholarly Communication and Academic Presses: Proceedings of the International Conference, 22 March 2001, University of Firenze, Italy*, edited by Anna Maria Tammaro, 2002
7. *Recenti acquisizioni nei disturbi del comportamento alimentare*, a cura di Alessandro Casini, Calogero Surrenti, 2003
8. *Proceedings of Physmod 2003 International Workshop on Physical Modelling of Flow and Dispersion Phenomena*, edited by Giampaolo Manfreda e Daniele Contini, 2003
9. *Public Administration, Competitiveness and Sustainable Development*, edited by Gregorio Arena, Mario P. Chiti, 2003
10. *Authority control: definizioni ed esperienze internazionali: atti del convegno internazionale, Firenze, 10-12 febbraio 2003*, a cura di Mauro Guerrini e Barbara B. Tillet; con la collaborazione di Lucia Sardo, 2003
11. *Le tesi di laurea nelle biblioteche di architettura*, a cura di Serena Sangiorgi, 2003
12. *Models and analysis of vocal emissions for biomedical applications: 3rd international workshop: december 10-12, 2003 : Firenze, Italy*, a cura di Claudia Manfredi, 2004
13. *Statistical Modelling. Proceedings of the 19th International Workshop on Statistical Modelling: Florence (Italy) 4-8 july, 2004*, edited by Annibale Biggeri, Emanuele Dreassi, Corrado Lagazio, Marco Marchi, 2004
14. *Studi per l'insegnamento delle lingue europee : atti della prima e seconda giornata di studio (Firenze, 2002-2003)*, a cura di Maria Carlota Nicolás Martínez, Scott Staton, 2004.
15. *L'Archivio E-prints dell'Università di Firenze: prospettive locali e nazionali. Atti del convegno (Firenze, 10 febbraio 2004)*, a cura di Patrizia Cotoneschi, 2004
16. *TRIZ Future Conference 2004. Florence, 3-5 November 2004*, edited by Gaetano Cascini, 2004
17. *Mobbing e modernità : la violenza morale sul lavoro osservata da diverse angolature per coglierne il senso, definirne i confini. Punti di vista a confronto. Atti del Convegno Firenze, 20 aprile 2004*, a cura di Aldo Mancuso, 2004
18. *Lo spazio sociale europeo. Atti del convegno internazionale di studi Fiesole (Firenze), 10-11 ottobre 2003*, a cura di Laura Leonardi, Antonio Varsori, 2005
19. *AIMETA 2005 Atti del XVII Congresso dell'Associazione Italiana di Meccanica Teorica e Applicata, Firenze, 11-15 settembre 2005*, a cura di Claudio Borri, Luca Facchini, Giorgio Federici, Mario Primicerio, 2005
20. *Giornata CEFTrain - CEFTrain Day. Conference Proceedings and Partner Contributions. Firenze, Italy, 7 may 2005*, edited by Elizabeth Guerin, 2005

**MODELS AND ANALYSIS  
OF VOCAL EMISSIONS  
FOR BIOMEDICAL APPLICATIONS**

**4th INTERNATIONAL WORKSHOP**

**October 29-31, 2005**

**Firenze, Italy**

**Edited by**

**Claudia Manfredi**

Firenze University Press

2005

Models and analysis of vocal emissions for biomedical applications :  
4th international workshop: october 29-31, 2005 : Firenze, Italy /  
edited by Claudia Manfredi. – Firenze : Firenze university press,  
2005.

(Atti, 21)

<http://digital.casalini.it/8884533201>

Stampa a richiesta disponibile su <http://epress.unifi.it>

ISBN 88-8453-320-1 (online)

ISBN 88-8453-319-8 (print)

612.78 (ed. 20)

Voce - Patologia medica

Responsibility for the contents rests entirely with the authors. The editors and the organising committee of the MAVEBA 2003 accept no responsibility for any errors, omissions, or views expressed in this publication.

No part of this publication can be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the permission of the editors. Permission is not required to copy abstracts of papers, on condition that a full reference of the source is given.

Cover: designed by CdC, Firenze, Italy.

© 2005 Firenze University Press

Università degli Studi di Firenze

Firenze University Press

Borgo Albizi, 28, 50122 Firenze, Italy

<http://epress.unifi.it/>

*Printed in Italy*



# MAVEBA 2005

## INTERNATIONAL PROGRAM COMMITTEE

S. Aguilera (ES)	A. Barney (UK)	D. Berckmans (BE)	P. Brusaglioni (I)
S. Cano Ortiz (CU)	R. Carlson (SE)	M. Clements (USA)	J. Doorn (AR)
U. Eysholdt (D)	O. Fujimura (JP)	H. Herzel (D)	D. Howard (UK)
A. Izvorski (PL)	M. Kob (D)	A. Krot (BY)	U. Laine (FI)
C. Larson (USA)	F. Locchi (I)	J. Lucero (BR)	C. Manfredi (I)
C. Marchesi (I)	W. Mende (D)	V. Misun (CZ)	C. Moore (UK)
X. Pelorson (F)	P. Perrier (F)	R. Ritchings (UK)	S. Ruffo (I)
O. Schindler (I)	A. Schuck (BR)	R. Shiavi (USA)	H. Shutte (NL)
J. Sundberg (SE)	J. Svec (CZ)	R. Tadeusiewicz (PL)	I. Titze (USA)
U. Uergens (D)	G. Valli (I)	K. Wermke (D)	W. Ziegler (D)

## LOCAL ORGANISING COMMITTEE

**C. Manfredi**, Dept. of Electronics and Telecommunications, Faculty of Engineering - Conference Chair  
**L. Bocchi**, Dept. of Electronics and Telecommunications, Faculty of Engineering  
**P. Brusaglioni**, Dept. of Physics, Faculty of Maths, Physics and Natural Sciences - Conference Chair  
**F. Locchi**, Dept. of Clinical Physiology, Faculty of Medicine.  
**C. Marchesi**, Dept. of Systems and Computer Science, Faculty of Engineering.  
**S. Ruffo**, Dept. of Energetics, Faculty of Engineering.  
**G. Valli**, Dept. of Electronics and Telecommunications, Faculty of Engineering

## SPONSORS

**Ente CRF** – Ente Cassa di Risparmio di Firenze



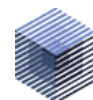
**ISCA** - International Speech and Communication Association



**AIIMB** - Associazione Italiana Ingegneria Medica e Biologica



**INFN** – Istituto Nazionale per la Fisica della Materia



# CONTENTS

Foreward.....	XI
---------------	----

## Special session on Voice pathology classification

C. Berry, T. Ritchings, “A comparative study of intelligent voice quality assessment using impedance and acoustic signals” .....	3-6
C. J. Moore , K. Manickam, N. Slevin, “Voicing recovery in males following radiotherapy for larynx cancer” .....	7-10
N. Sáenz-Lechón, J.I. Godino-Llorente, V. Osma-Ruiz, P. Gómez-Vilda, S. Aguilera-Navarro, “A methodology to evaluate pathological voice detection systems” .....	11-14
N. Sáenz-Lechón, J.I. Godino-Llorente, V. Osma-Ruiz, P. Gómez-Vilda, S. Aguilera-Navarro, “Effects of MP3 encoding on voice pathology detection: results with MFCC parameters” .....	15-18
J. Schoentgen, “Towards a classification of phonatory features of disordered voices” .....	19-22
W. Sheta, T. Ritchings, C. Berry, “Intelligent voice quality assessment post-treatment using genetic programming” .....	23-26

## Voice recovering / enhancement

H. Kuwabara, “A method for changing speech quality and its application to pathological voices” .....	29-32
R. Pietruch, A. Grzanka, W. Konopka, M. Michalska, “Methods for formant extraction in speech of patients after total laryngectomy” .....	33-36
C. Manfredi, C. Marino, F. Dori, E. Iadanza, ”Optimised GSVD for dysphonic voice quality enhancement” .....	37-40

## Special session on Physical and mechanical models and devices

T. Vampola, J. Horacek, J. Vesely, J. Vokral, “Modelling of influence of velopharyngeal insufficiency on phonation of vowel /a/” .....	43-46
R. Zaccarelli, C. Elemans, W.T. Fitch, H. Herzel, “Two-mass models of the bird Syrinx” .....	47-50
N. Ruty, J. Cisonni, X. Pelorson, P. Perrier, P. Badin, A. Van Hirtum, “A physical model for articulatory speech synthesis. theoretical and numerical principles” .....	51-54
F. Avanzini, S. Maratea, C. Drioli, “Physiological control of low-dimensional glottal models with applications to voice source parameter matching” .....	55-58

P. Gómez, C. Lázaro, R. Fernández, A. Nieto, J.I. Godino, R. Martínez, F. Díaz, A. Álvarez, K. Murphy, V. Nieto, V. Rodellar, F.J. Fernández-Camacho, “ <i>Using biomechanical parameter estimates in voice pathology detection</i> ” .....	59-62
R. Orr, B. Cranen, “ <i>The effect of the flow mask on phonation</i> ” .....	63-66
Ch. Jeannin, P. Perrier, Y. Payan, A. Dittmar, B. Grosgeat, “ <i>A non-invasive device to measure mechanical interaction between tongue, palate and teeth during speech production</i> ” .....	67-70

### Poster session

P. Alku, J. Horacek, M. Airas, A.M. Laukkanen, “ <i>Assessment of glottal inverse filtering by using aeroelastic modelling of phonation and fe modelling of vocal tract</i> ” .....	73-76
L. Bocchi, S. Bianchi, C. Manfredi, N. Migali, G. Cantarella, “ <i>Quantitative analysis of videokymographic images and audio signals in dysphonia</i> ” .....	(Paper not available)
P. Brusciagioni, “ <i>A note on jitter estimation</i> ” .....	(Paper not available)
M. Guarino, A. Costa, S. Patelli, M. Silva, D. Berckmans, “ <i>Cough analysis and classification by labelling sound in swine respiratory disease</i> ” .....	77-80
J. Hanquinet, F. Grenez, J. Schoentgen, “ <i>Wave-morphing in the framework of a glottal pulse model</i> ” .....	81-83
T. Kitamura, P. Mokhtari, H. Takemoto, “ <i>Changes of vocal tract shape and area function by F0 shift</i> ” .....	85-88
M. Kob, J. Stoffers, Ch. Neuschaefer-Rube, “ <i>Comparison of LPC analysis and impedance vocal tract measurements</i> ” .....	89-91
C. Manfredi, B. Maraschi, A. Berlusconi, G. Cantarella, “ <i>Objective pre- and post-surgical voice analysis</i> ” .....	(Paper not available)
V. Misun, “ <i>Source voice characteristics of the artificial vocal folds</i> ” .....	93-96
T. Orzechowski, A. Izvorski, I. Gatkowska, M. Rudzinska, “ <i>Automatic classification of voice disorders in course of neurodegenerative disease</i> ” .....	97-100
M. Sovilj, T. Adamovic, M. Subotic, N. Stevovic, “ <i>Newborn’s cry from risk and normal pregnancies</i> ” .....	101-104
K. Manickam, H. Li, “ <i>Complexity analysis of normal and deaf infant cry acoustic waves</i> ” .....	105-108

### Special session on Methods for voice measurements

A. Fourcin, “ <i>Clinical voice measurement using EGG/LX signals</i> ” .....	111-114
R. Shrivastav, “ <i>From vocal quality measurement to perception</i> ” .....	115-118

J.G. Svec, F. Sram, M. Fric, Q. Qiu, H.K. Schutte, “ <i>What can be seen in videokymographic images?</i> ” .....	119
S. Bianchi, L. Bocchi, C. Manfredi, G. Cantarella, N. Migali, “ <i>Objective vocal fold vibration assessment from videokymographic images</i> ” .....	121-124
S. Mantha, L. Mongeau, T. Siegmund, “ <i>Dynamic digital image correlation of a dynamic physical model of the vocal folds</i> ” .....	125-128
S. Ciecwiwa, D. Deliyiski, T. Zielinski, “ <i>Fast FFT-based motion compensation for laryngeal high-speed videoendoscopy</i> ” .....	129-132
H.S. Shaw, D.D. Deliyiski, “ <i>Vertical motion during modal and pressed phonation: magnitude and symmetry</i> ” .....	133-136
H.S. Shaw, D.D. Deliyiski, “ <i>Mucosal wave magnitude: presence, extent, and symmetry in normophonic speakers</i> ” .....	137-140
S.T. Takano, K.K. Kinoshita, K.H. Honda, “ <i>Measurement of cricothyroid articulation using high-resolution MRI and 3D pattern matching</i> ” .....	141-144

### **Voice modelling and analysis**

H.J. Fell, J. MacAuslan, “ <i>Vocalisation analysis tools</i> ” .....	147-150
P. Svancara, J. Horacek, “ <i>Numerical modelling of effect of tonsillectomy on production of Czech vowels /a/ and /i/</i> ” .....	151-154
A. Kacha, F. Grenez, J. Schoentgen, “ <i>Generalized variogram analysis of vocal dysperiodicities in connected speech</i> ” .....	155-158
T. Wurzbacher, R. Schwarz, H. Toy, U. Eysholdt, J. Lohscheller, “ <i>Modelling of non-stationary phonation for classification of vocal folds vibrations</i> ” .....	159-162
F.M. Martinez, J.C. Goddard, A.M. Martinez, “ <i>Analysis of Spanish synthesized speech signals using spectral and basis pursuit representations</i> ” .....	163-166
F.R. Drepper, “ <i>Voiced excitation as entrained primary response of a reconstructed glottal master oscillator</i> ” .....	167-170
J.J. Turunen, T. Lipping, J.T. Tantt, “ <i>Speech analysis using Higuchi fractal dimension</i> ” .....	171-174

### **Special session on Neurological dysfunctions**

R. Shiavi, “ <i>Quantitative and experimental approaches for investigating neurological/psychological dysfunction</i> ” .....	(Paper not available)
L. Cnockaert, J. Shoentgen, P. Auzou, C. Ozsancak, F. Grenez, “ <i>Effect of Parkinson’s disease on vocal tremor</i> ” .....	177-180



K.S. Subari, D.M. Wilkes, R. Shiavi, S. Silverman, M. Silverman, “ <i>Evaluation of speaker normalization for suicidality assessment</i> ” .....	181-184
--	---------

### **Infant cry – Singing voice**

K. Wermke, W. Mende, A. Kempf, C. Manfredi, P. Brusaglioni, A. Stellzig-Eisenhauer, “ <i>Interaction patterns between melodies and resonance frequencies in infants’ pre-speech utterances</i> ” .....	187-190
--	---------

R. Nicollas, J. Giordano, L. Francius, J. Vicente, Y. Burtschell, M. Medale, B. Nazarian, M. Roth, M. Ouaknine, A. Giovanni, “ <i>Aerodynamical model of human newborn larynx: an approach of the first cry</i> ” .....	191-193
---	---------

S. Adachi, J. Yu, “ <i>High-pitched voice simulation using a two-dimensional vocal fold model</i> ” .....	195-198
---	---------

J. Sundberg, “ <i>Effect of vocal loudness variation on the voice source</i> ” .....	199
--	-----

N. Amir, O. Amir, O. Michaeli, “ <i>Assessing vibrato quality of singing students</i> ” .....	201-203
---	---------

### **Non-human sounds**

M. Gamba, C. Giacoma, “ <i>Vocal production mechanism in ruffed lemurs: a prosimian model for the basis of primate phonation</i> ” .....	207-210
--	---------

J.-M. Aerts, P. Jans, D. Halloy, P. Gustin, D. Berckmans, “ <i>Labelling of cough data from pigs for on-line disease monitoring by sound analysis</i> ” .....	211-214
---	---------

Author Index.....	215-216
-------------------	---------



# MAVEBA 2005

## FOREWARD

Welcome to the 4<sup>th</sup> International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, MAVeBA 2005, 29-31 October 2005, Firenze, Italy.

In the light of previous editions, held in 1999, 2001, and 2003 respectively, all in Firenze, the Workshop aims at investigating into main aspects of voice modelling and analysis, ranging from fundamental research to all kinds of biomedical applications and related established and advanced technologies. It offers the participants an interdisciplinary platform for presenting and discussing new knowledge both in the field of models and analysis of speech signals and in that of emerging imaging techniques.

Contacts between specialists active in research and industrial developments could take advantage from the Workshop structure, comprising both Special Sessions devoted to a set of relevant topics, and standard Sessions, covering a wide area of voice analysis research, for biomedical applications.

Four Special Sessions were organised and co-ordinated by specialists in the field, collecting contributions about new and emerging techniques. Each Session is introduced by a review paper, presenting the state-of-the-art in the field, pointing out present knowledge, limitations and future directions. The selected topics are:

1. Voice pathology classification
2. Physical and mechanical models and devices
3. Methods for voice measurements
4. Neurological dysfunctions

As for regular Sessions, the relevant topics are: voice recovering, enhancement of voice quality during rehabilitation and after surgery, voice modelling and analysis of vocal emissions, newborn and infant cry analysis, singing voice. A short Session is also devoted to non-human sounds, and their possible relationships to humans.

All the papers collected in this book of Proceedings are of high scientific level, as they were reviewed by at least two anonymous referees, and cover the most relevant fields of research in voice signals and images analysis. We would like to thank the members of the organising committee and all the reviewers, who gave freely of their time to assess the highly disparate work of the workshop, helping in improving the quality of the papers.

We have also benefited from the efforts of the administrative staff within our University: office for Research and International Relations, Logistic office, and the staff of the Faculty of Engineering and of the Department of Electronics and Telecommunications, that devoted a lot of time and efforts to make this workshop a successful one. Special thanks to our University Orchestra and Chorus, and to the members of “Capriccio Armonico” dancing group for their generous participation.

Finally, our gratitude goes to the supporters and sponsors, who contribute much to the success of the MAVeBA workshop.

Dott. Claudia Manfredi  
Conference Chair

Prof. Piero Bruscazioni  
Conference Chair

# OPTIMISED GSVD FOR DYSPHONIC VOICE QUALITY ENHANCEMENT

Claudia Manfredi, Cloe Marino, Fabrizio Dori, Ernesto Iadanza

Department of Electronics and Telecommunications  
Università degli Studi di Firenze, Via S.Marta 3, 50139 Firenze, Italy  
[manfredi@det.unifi.it](mailto:manfredi@det.unifi.it)

**Abstract:** This paper concerns the problem of enhancing voice quality for people suffering from dysphonia, caused by airflow turbulence in the vocal tract, for irregular vocal folds vibration.

A generalized subspace approach is proposed for enhancement of speech corrupted by additive noise, regardless of whether it is white or not. The clean signal is estimated by nulling the signal components in the noise subspace and retaining the components in the signal subspace. Two approaches are compared, taking into account both signal and noise, or signal only, eigenvalues. An optimised adaptive comb filter is applied first, to reduce noise between harmonics. Objective voice quality measures demonstrate improvements in voice quality when tested with sustained vowels or words corrupted with “hoarseness noise”. The intention is to provide users (disabled people, as well as clinicians) with a device allowing intelligible and effortless speech for dysphonics, and useful information concerning possible functional recovering. This will be of use to people in social situations where they interact with non-familiar communication partners, such as at work, and in everyday life.

**Keywords:** hoarseness, voice denoising, GSVD, comb filtering, voice quality, pitch, noise, formants.

## I. INTRODUCTION

Signal subspace methods are used frequently for denoising in speech processing, mainly with speech communication [1], [2]. Until now, few results are available concerning their application for voice quality enhancement in the biomedical field [3]. In this paper, the objective of noise reduction is to improve noisy signals due to irregular vocal folds vibration. This problem is of great concern, for rehabilitation and from the assistive technology point of view. Commonly, surgical and/or pharmacological treatments allow restoring voice quality, with patient’s recovering to an acceptable or even excellent level. However, sometimes patients can only partly recover, with heavy implications on their quality of life.

The idea behind subspace methods is to project the noisy signal onto two subspaces: the signal subspace (since the signal dominates this subspace), and the noise subspace. The noise subspace contains signals from the noise

process only, hence an estimate of the clean signal can be made by removing or nulling the components of the signal in the noise subspace and retaining only the components of the signal in the signal subspace. The decomposition of the space into two subspaces can be done using either the singular value decomposition (SVD) [4], [5] or the Quotient SVD (QSVD) or GSVD [1],[6],[11]. Though computationally expensive, GSVD was found robust and effective in reducing noise due to turbulences in the vocal tract, which is typically coloured. GSVD is implemented here with two choices for separating the signal and the noise subspaces, to compare performance. Specifically, the first choice is based on classical GSVD, where both the signal and the noise subspace eigenvalues are used for filtering [6]. The second one corresponds to retaining the signal subspace eigenvalues only [1].

An adaptive comb filter is applied first, as it was shown to significantly reduce noise between the harmonics in the spectrum. The comb filter is optimised, in the sense that it is applied on windows whose length varies according to varying pitch.

Real data coming from dysphonic subjects are successfully denoised with the proposed approaches.

## II. MATERIALS AND METHODS

Firstly, optimised adaptive comb filtering is performed on data windows of varying length, obtained with a new two-step robust adaptive pitch estimation technique [7]. The essence of comb filtering is to build a filter that passes the harmonics of the noisy speech signal  $y$ , while rejecting noise frequency components between the harmonics [8],[9]. Ideally, spacing between each “tooth” in the comb filter should correspond to  $F_0(1/T_0)$  in Hz, which is often highly unstable in pathological voices. The proposed comb filter, based on an adaptive two-step pitch estimator, is capable to adapt to fast pitch variations and successfully reduces noise as evaluated by an adaptive implementation of the Normalised Noise Energy technique (ANNE) [7], thus being suited as a pre-filtering step. The filter that has been used in this paper has a Hamming window shape, which is obtained from the following equation (with  $K=3$ ):

$$a(i) = \frac{0.54 + 0.46 \cos(2\pi i / 2K + 1)}{\sum_{i=-K}^K 0.54 + 0.46 \cos(2\pi i / 2K + 1)} \quad (1)$$

This step is followed by Generalised Singular Value Decomposition (GSVD) of signal and noise matrices, whose entries are suitably organised, as shown in eq. (4). GSVD-based voice denoising aims at diminishing the uncorrelated and added noise from the voice signal, whether it is white or not. The noisy signal  $y$  at time instant  $t$ ,  $y_t$ , can be expressed as:

$$y_t = d_t + n_t \quad (2)$$

Where  $d$ =clean signal,  $n$ =(coloured) noise. The goal is to estimate  $d$  from  $y$ . The noisy signal is segmented into frames  $y_i$ ,  $i=1, 2, \dots$ , of varying length  $M_i$ , obtained according to the previously cited robust adaptive pitch estimation procedure. The GSVD amounts to finding a non-singular matrix  $X$  and two orthogonal matrices  $U$ ,  $V$  of compatible dimensions, which simultaneously transform both the Hankel noisy speech matrix  $H_y$  and the noise matrix  $H_n$  into nonnegative diagonal form matrices  $C$  and  $S$  such as:

$$\begin{aligned} U^T H_y X &= C = \text{diag}(c_1, \dots, c_k), c_1 \geq c_2 \geq \dots \geq c_k \\ V^T H_n X &= S = \text{diag}(s_1, \dots, s_k), s_k \geq s_{k-1} \geq \dots \geq s_1 \\ C^T C + S^T S &= I_k \end{aligned} \quad (3)$$

Where  $L+K=M+1$ ,  $K < L$ . The  $H_y$  matrix has the form:

$$H_y = \begin{bmatrix} y_0 & y_1 & \dots & y_{K-1} \\ y_1 & y_2 & \dots & y_K \\ \vdots & \vdots & \dots & \vdots \\ y_{L-1} & y_L & \dots & y_{M-1} \end{bmatrix} \quad (4)$$

Similarly for  $H_n$ .

The values  $c_1/s_1 \geq c_2/s_2 \geq \dots \geq c_k/s_k$  are referred to as the generalised singular values of  $H_y$  and  $H_n$ . Notice that one can choose to work with Toeplitz matrices instead of Hankel matrices. There are no fundamental differences between the two approaches.

It was shown [1], [2], [6], [11] that the filtered signal can be obtained either from the matrix:

$$H_y^p = U \begin{bmatrix} C_p S_p^{-1} & 0 \\ 0 & 0 \end{bmatrix} X^{-1} \quad (5)$$

or from the matrix:

$$H_y^p = U \begin{bmatrix} C_p & 0 \\ 0 & 0 \end{bmatrix} X^{-1} \quad (6)$$

where  $U$  and  $X$  are as in eq. (3) and  $C_p = \text{diag}(c_1, \dots, c_p)$ ,  $S_p = \text{diag}(s_1, \dots, s_p)$ , are sub-matrices of  $C$  and  $S$  respectively and  $p$  is the signal subspace dimension. Eq. (5) corresponds to classical GSVD, where both the signal and the noise subspace eigenvalues are used for filtering, and will be referred to as GSVD in what follows. Eq. (6) corresponds to retaining the signal subspace eigenvalues only, and will be referred to as OSV (Only Signal Values). Two problems were encountered with GSVD, i.e. the choice of the noise covariance matrix and that of the signal subspace dimension  $p$ . Commonly, in speech communication settings, the noise covariance matrix is computed using noise samples collected during speech-absent frames. To deal with the problem under study, different choices were tested. Among them, one takes

into account the signal noisy component as obtained from a preliminary SVD decomposition of the signal under study: the noise subspace is reconstructed and used to fill matrix  $H_n$ . While giving almost good results, this choice was disregarded, due to both the larger computational load and to better results obtained with the following approach: on each signal frame of varying length, an AutoRegressive (AR) model is identified, and the model residuals are evaluated. The residual variance is then used to construct the diagonal matrix  $S$  of eq. (3).

The second problem is the optimal choice of the number  $p$  of retained singular values for denoised signal reconstruction. Classical order selection criteria were applied to GSVD, such as AIC, MDL [9], as well as a new criterion named DME [10], but best results were obtained with  $p=2$ . It will be named as GSVD<sub>fix</sub> in what follows. As for OSV,  $p$  was chosen such as [1]:

$$c_p > s_p \text{ and } c_{p+1} < s_{p+1} \quad (7)$$

This was in fact the choice that gave the best results.

Finally, three objective indexes are defined, closely related to the signal characteristics. A frequency threshold value  $f_{th}=4\text{kHz}$  is defined, based on the usual range for voiced sounds (first four formants) in adults, as well as on experimental results obtained from threshold tuning in a dataset of voiced and unvoiced sounds. The subscript “non-filt” refers to the original signal, while “filt” refers to the denoised signal:

$$\text{PSD}_{\text{low}} = 10 \log_{10} \frac{\text{PSD}_{\text{non-filt}}(f \leq f_{th})}{\text{PSD}_{\text{filt}}(f \leq f_{th})} \quad (8)$$

measures the ratio of the PSDs evaluated on the “harmonic range”;

$$\text{PSD}_{\text{high}} = 10 \log_{10} \frac{\text{PSD}_{\text{non-filt}}(f \geq f_{th})}{\text{PSD}_{\text{filt}}(f \geq f_{th})} \quad (9)$$

is the ratio of the PSDs, evaluated on the “noise range”. A

$$\text{QER} = 10 \log_{10} \frac{\sum_{n=1}^M y^2(n)}{\sum_{n=1}^M (y(n) - y_{\text{filt}}(n))^2} \quad (10)$$

good denoising procedure should give  $\text{PSD}_{\text{low}}$  values near to zero (no loss of harmonic power), but high  $\text{PSD}_{\text{high}}$  values (loss of power due to noise).

Finally, a measure of the denoising effectiveness (quality enhancement ratio, QER) is defined as:

QER is thus the ratio between the signal energy and that of the removed noise.  $\text{QER} > 0$  corresponds to good denoising [10].

### III. RESULTS

A set of about 20 voice signals (word /aiuole/) coming from adult male patients were analysed with the proposed approach. All patients underwent surgical removal of T1A glottis cancer, by means of laser or lancet technique. Perceptual evaluation with GIRBAS scale showed good recovering, however, residual hoarseness was found in