3rd International Workshop MAVEBA 2003, 69-72

HOARSE VOICE DENOISING FOR REAL-TIME DSP IMPLEMENTATION: CONTINUOUS SPEECH ASSESSMENT

E. Iadanza[1], F. Dori[1], C. Manfredi[1], S. Dubini[1]
[1]Department of Electronics and Telecommunications, Università degli Studi di Firenze, Firenze, Italy

***Abstract*: Voice hoarseness is mainly related to airflow turbulence in the vocal tract. It can be due to vocal fold paralysis, polyps, cordectomisation or other dysfunction, which alter regular speech production, and is commonly treated as a noise component in the speech signal. A denoising approach is proposed, based on low-order singular value decomposition (SVD) of matrices whose entries come from sampled speech data frames, properly organised. A prototype DSP board implementing the procedure was developed. Objective quality indexes are proposed, showing the results achieved with the proposed method both on vowel and consonantal sentences.**

***Keywords:* SVD, hoarse voice, DSP, continuous speech, real-time**

I. INTRODUCTION

This paper deals with the problem of enhancing voice quality for people suffering from dysphonia. This can be due to vocal fold paralysis, cordectomisation or other dysfunction, which alter regular speech production and commonly cause more efforts to be used in speaking than for healthy people. Objective speech quality measures are reliable, easy to implement and have been shown to be good predictors of subjective quality [7], [16]. The main goal of the system presented here is to realise a mobile hardware/software system for real-time voice denoising, to obtain a more intelligible speech with small effort. The method is based on the singular value decomposition (SVD) of matrices whose entries come from sampled speech data frames, properly organised [1]. SVD is widely used for speech enhancement, mainly to improve the performance of speech communication systems in a noisy environment [2], [3], [4]. For the present application, a fixed two-dimensional signal subspace dimension was found sufficient for data filtering, thus allowing real-time implementation. Objective quality measures (PSD ratios, SNR) are defined and evaluated, in order to assess enhancement of voice and compare results. The proposed approach was implemented on a DSP board, by means of properly optimised C and Assembler code. Thus, a simple portable device could be realised, as an aid for dysphonic speakers for diminishing effort in speaking, which is closely related to social problems due to awkwardness of voice.

II. DENOISING WITH SVD

The SVD is a numerically reliable and robust means for estimating the space of clean data (signal subspace) from the white noise corrupted data, and is thus particularly suited for speech denoising [1], [5], [6], [7], [8]. Despite its simplicity, the SVD approach was found effective in increasing voice quality. Extensive simulations were performed and detailed results are reported in [9], [10]. This paper aims at testing the method on continuous speech, to evaluate its performance on consonantal sounds mixed to vocalic ones. Moreover, in order to measure performance, some simple objective quality indexes will be introduced and evaluated.

III. QUALITY MEASURES

Extensive research has been carried out in developing both subjective and objective tests to ascertain quality, but few results are available as far as correlation among them is concerned [16]. In the following, some indexes are proposed, closely related to the signal characteristics. In this work it is assumed that "harmonic" range means frequencies below $f_{th} = 4kHz$, while "noise" range indicates frequencies over this threshold. This threshold is an empiric choice based on analysis of various speech signals; we are currently tuning it using a wider dataset. The subscript "non-filt" refers to the original signal, while "filt" refers to the SVD-filtered signal. The simplest measure is:

$$\mathrm{PSD} = 10\log_{10}\frac{\mathrm{PSD}_{\mathrm{non-filt}}}{\mathrm{PSD}_{\mathrm{filt}}} \qquad (1)$$

representing the ratio of the PSDs, evaluated on the whole frequency range;

$$\mathrm{PSD}_{\mathrm{low}} = 10\log_{10}\frac{\mathrm{PSD}_{\mathrm{non-filt}}(\mathrm{f} \le 4\mathrm{kHz})}{\mathrm{PSD}_{\mathrm{filt}}(\mathrm{f} \le 4\mathrm{kHz})} \qquad (2)$$

measures the ratio of the PSDs evaluated on the "harmonic" range, while

$$\mathrm{PSD}_{\mathrm{high}} = 10\log_{10}\frac{\mathrm{PSD}_{\mathrm{non-filt}}(\mathrm{f} \ge 4\mathrm{kHz})}{\mathrm{PSD}_{\mathrm{filt}}(\mathrm{f} \ge 4\mathrm{kHz})} \qquad (3)$$

is the ratio of the PSDs, evaluated on the "noise" range.

A good denoising procedure should give PSD and $\mathrm{PSD}_{\mathrm{low}}$ values around zero (no loss of power), but high

PSD_{high} values (loss of power due to noise). Finally,

$$SNR = 10\log_{10} \frac{\sum_{n=1}^{M} y^2(n)}{\sum_{n=1}^{M} (y(n) - y_{filt}(n))^2} \quad (4)$$

where: y(n) = noisy signal sample at time n, $y_{filt}(n)$ = filtered signal sample at time n.

Notice that PSD_{low} and SNR have good correlates with NHR [16] and the GIRBAS scale, while being simple and reliable at a very low computational cost. This point will be further exploited in future work.

IV. Experimental results

The denoising procedure was applied here to real data. These concern hoarse pathological voices, coming from adult male subjects that underwent partial cordectomisation, due to T1A glottis cancer. Patients were asked to pronounce the Italian word /aiuole/ (flowerbeds), which is composed of the five principal vowels. This choice is due to the clinical interest in evaluating the effort in speaking made by patients, for surgical and rehabilitation purposes.Besides, the method has been also tested on a pathologic subject pronouncing a 12 sec. sentence taken from Kay Elemetrics disordered voice database, developed by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab.

The results from SVD filtering procedure applied on the word /aiuole/ were compared to those coming from the complete phrase, by means of the quality indexes described in sect.3 in order to evaluate the method's performance also on non-vocal sounds and silence.

Fig. 1 shows the results relative to one subject (lancet operated) pronouncing the word /aiuole/. The approach lowers the PSD on the whole frequency range (PSD=0.02 dB), and especially on the low frequency range (PSD_{low}=-0.004 dB). This corresponds to a good voice level at the output of the filtering chain. Good value is also found on the high frequency region (PSD_{high}=14.6 dB), and correspondingly a SNR value near to 16 dB (SNR=16.4 dB). Fig. 1 shows the spectrogram of the unprocessed signal (upper plot), as compared to that obtained from the SVD filtering chain (lower plot). For clearness, the frequency range is limited to a maximum of 6 kHz. The lower plot confirms the good denoising properties of the proposed procedures, as the noise level is largely reduced above 4 kHz. As already said, denoising with the proposed SVD approach preserves the temporal and spectral characteristics of the original signal, thus providing a filtered voice of better quality, without distorting effects. Fig. 2-3 plot the results obtained for a 12 sec sentence (hence, not just vowel sounds).

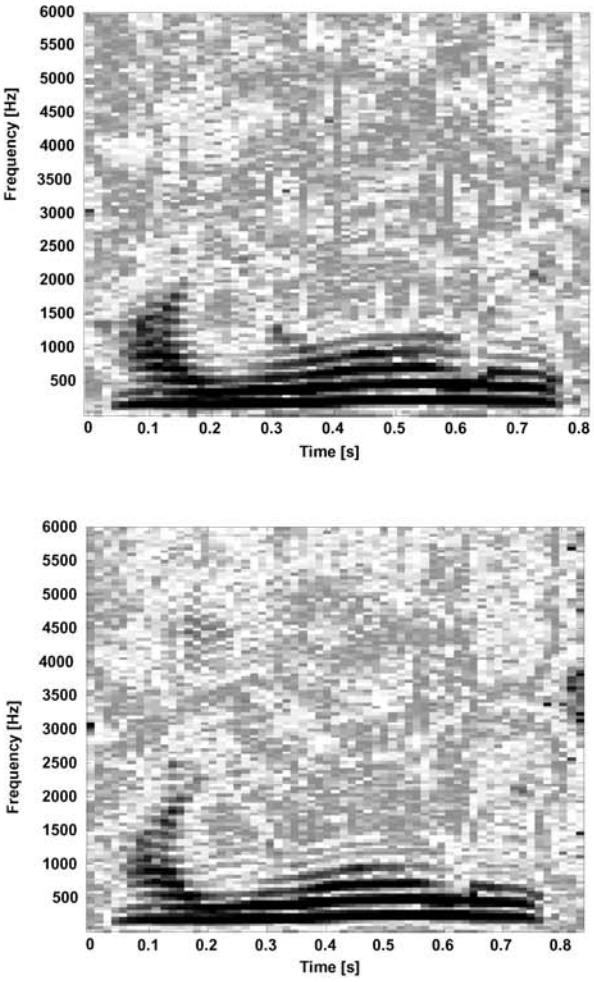

Figure 1 – Spectrogram of the signal before denoising (lower), after denoising (upper).

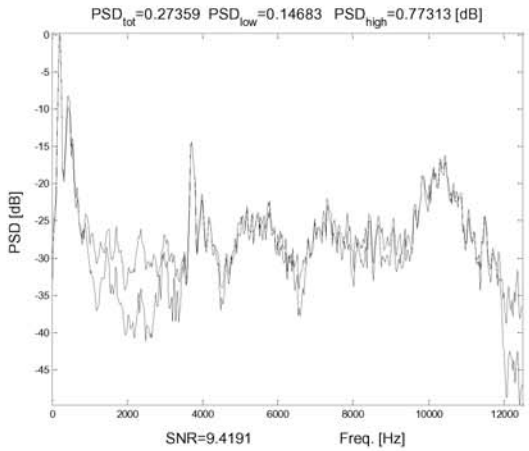

Figure 2 – Comparison of PSD plots for non-filtered (solid line) and for the filtered sentence (dotted line)

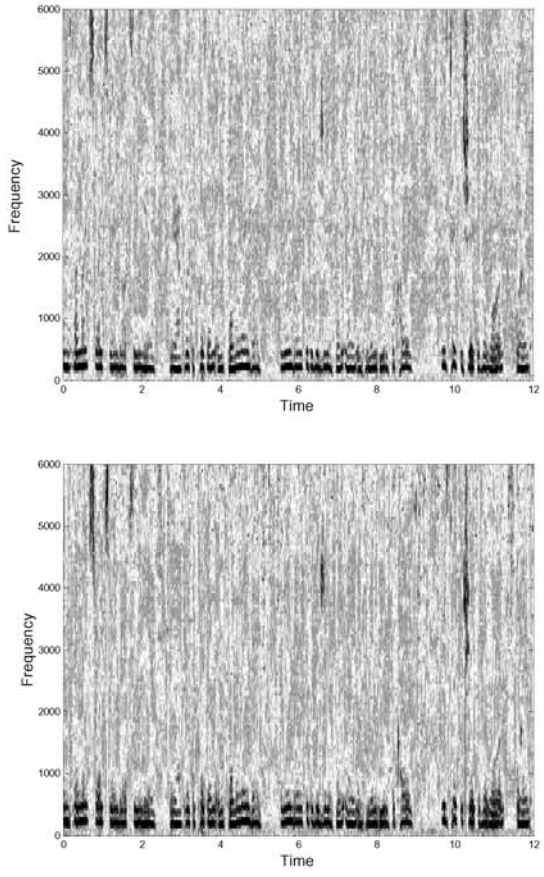

Figure 3 – Spectrogram of the naturally speaking signal before denoising (upper), after denoising (lower).

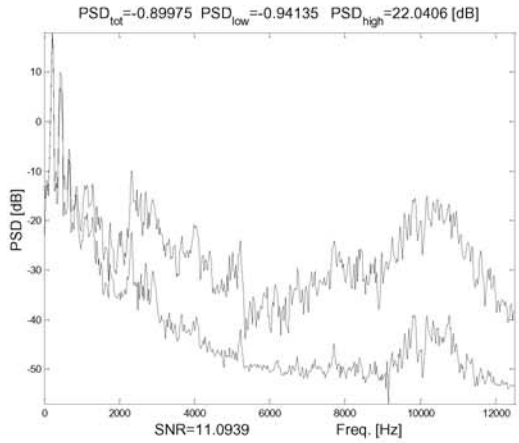

Figure 4 –PSD plots for non-filtered (solid line) and for the filtered naturally speaking signal (dotted line) (/rainbow/).

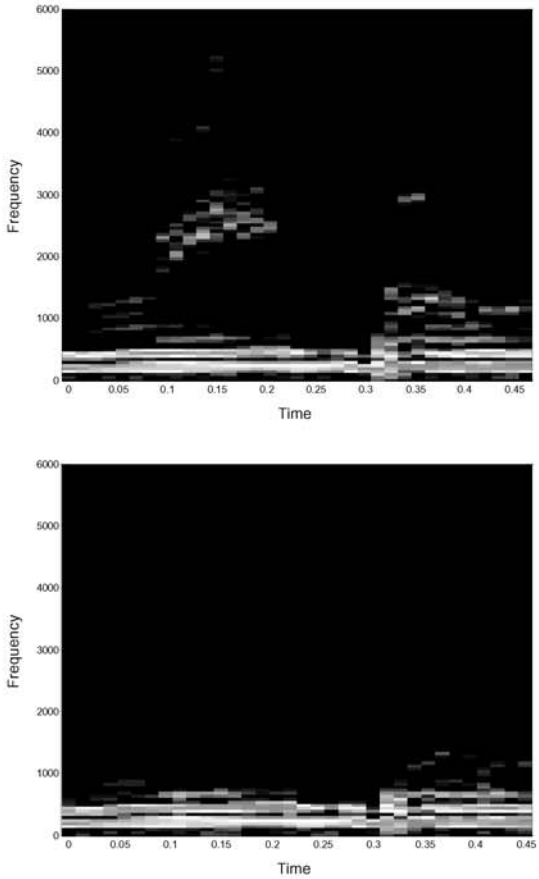

Figure 5 – Spectrogram of the naturally speaking signal before (upper), after denoising (lower) (/rainbow/). Colormap rescaled to fit signal dynamics.

Fig. 2 shows the PSD evaluated for the non-filtered signal (solid line) and for the signal filtered with the proposed method (dashed line). Low PSD values are found both for the PSD on the whole frequency range (PSD=0.274 dB), and for the low and high frequency ranges (PSD_{low}=0.147 dB; PSD_{high}=0.773 dB), and correspondingly a SNR value near to 10 dB (SNR= 9.4191 dB). The results basically correspond to close power values in output and input signals. This means that the system correctly doesn't cut informative signals in unvoiced sounds, even in frequencies above 4 kHz, while it shows strong denoising capabilities in noisy signals. Actually Fig. 3 highlights that noise level is widely reduced above 4 kHz while the so called harmonic range is left nearly unchanged. Specifically, the SVD approach allows lowering the noise component especially with voiced sounds, where the informative content is most in the harmonic range, while has negligible effect for unvoiced sounds, where the informative content is shared out both in the harmonic and in the noise range. Figs. 4-5 point out this aspect. The word /rainbow/ (prevalence of

vocalic sounds) is taken out from the whole sentence, giving good results. Fig. 4 shows very good values both for the PSD on the whole frequency range (PSD=0.9 dB), and especially for the low and high frequency ranges (PSD_{low}=0.941 dB; PSD_{high}=22.041 dB), and correspondingly a SNR value near to 11 dB (SNR= 11.094 dB). Fig. 5 confirms these results, being comparable to those in Fig.1.

V. HARDWARE/SOFTWARE IMPLEMENTATION

The software development tool integrates a C compiler/linker and the DSP/BIOS firmware for implementing a basic kernel with run-time services [11]. The SVD algorithm is implemented by means of a two-step procedure: first, the data matrix A is bi-diagonalised applying a sequence of Householder reflections; second, A is made diagonal using a modified QR algorithm [12-16]. The criteria adopted to implement the hardware platform are:

- High processing performance.
- Low power consumption/Low cost.

The board is supplied with analog front-end, capable to accept the audio signal as input and to furnish the output processed signal at the output stereo jack. The DSP-based board allows to process signals in the 0-48kHz bandwidth. For further details see [17]. The developed hardware was tested with real data in order to reach the real-time processing requirements.

VI. FINAL REMARKS

A simple approach for enhancing voice quality in dysphonic subjects is proposed. The method applies SVD for data filtering, separating the clean signal from its noisy component. The denoised signal is reconstructed along the directions spanned by the principal eigenvectors of the signal subspace. For filtering purposes, the best choice was found that of picking only the two dominant eigenvalues, thus resulting in a low-cost procedure, suitable for on-line implementation on a DSP board. The tests with whole sentences, as well as voiced sounds only, show that this method is suitable both for sustained vowels analysis and for portable application devices.

VII. REFERENCES

[1] Rao B D, Arun K S., "Model based processing of signals: a state space approach", *Proc. IEEE*, vol.80, 1992, pp. 283-309.

[2] Asano F, Hayamizu S, Yamada T, Nakamura S., "Speech enhancement based on the subspace method", IEEE *Trans. Speech Audio Proc.*, vol.8, 2000, pp.497-507.

[3] Ephraim Y, "Statistical model-based speech enhancement systems", *Proc. IEEE*, vol.80, 1992, pp.1526-1558.

[4] Ephraim Y, Van Trees H L., "A signal subspace approach for speech enhancement", *IEEE Trans. Speech Audio Proc.*, vol.3, 1995, pp.251-266.

[5] Klemma V C, Laub AJ. "The singular value decomposition: its computation and some applications", *IEEE Trans. Automat. Control*, vol. 25, 1980, pp.164-176.

[6] Marple S L., "Digital spectral analysis with applications", Prentice Hall, Englewood Cliffs, NJ, 1987.

[7] Deller J R, Proakis J G, Hansen J H L., "Discrete-time Processing of Speech Signals", Maxwell McMillan, New York, 1993.

[8] Manfredi C., "Adaptive noise energy estimation in pathological speech signals", *IEEE Trans. Biomed. Eng.*, vol.47, 2000, pp.1538-1542.

[9] Manfredi C., D'Aniello M., Bruscaglioni P., "A simple subspace approach for speech denoising", *Logopedics Phoniatrics Vocology*, vol.26, p.179-192, 2001.

[10] Manfredi C., Landini L., Faita F., Gemignani V. SVD-based portable device for real-time hoarse voice denoising. Proc. Int. Conf. Digital Signal Processing, Santorini, GR, 2002, pp. 857-860.

[11] Hirano M., "Psycho-acoustic evaluation of voice", In: Hirano M. Clinical examination of voice, Springer-Verlag, New York, 1981.

[12] Golub G.H., Van Loan C.F., "Matrix Computations", 2nd Ed., Johns Hopkins University Press, 1989.

[13] Forsythe G.E., Malcolm M.A., Moler C.B., "Computer methods for mathematical computations", Prentice-Hall, 1977.

[14] Stoer J., Bulirsch R., "Introduction to numerical analysis", Springer-Verlag, 1980.

[15] Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. T., "Numerical recipes in C – The art of scientific", Cambridge University Press, 1988.

[16] Dejonckere P H, Remacle M, Fresnel-Elbaz F, Woisard V, Crevier-Buchman L, Millet B, "Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements", *Rev. Laryngol. Otol. Rhinol*., vol.117, n.3, 1996, pp.219-224.

[17] Manfredi C., Dori F., Iadanza E., "Improvement in hoarse voice denoising for real-time DSP implementation", Proc. Int. Conf. Voice quality: functions, analysis and synthesis, Geneve, 2003