

Objective and Subjective Comparison of Audio Restoration Systems

Sergio Canazza (*), *Guido Corradu* (*), *Giovanni De Poli* (*),
Gian Antonio Mian (*)

(*) Dept. of Electronics and Informatics - University of Padova
V. Gradenigo 6/a - 35100 Pd - Italy
{canazza, depoli, mian}@dei.unipd.it

ABSTRACT

This paper presents the results of a perceptual test to evaluate the quality of different audio restoration algorithms based on Short Time Spectral Attenuation method and on Kalman filter. The protocol by International Communication Union (ITU.R P.800) was used. A set of audio stimuli played with different instruments (i.e. different bandwidths, transients, and timbre) was selected and different impulsive and broadband noises were added. Commercial software and proprietary algorithms were used to restore the corrupted stimuli. Twelve subjects were asked to compare the restored and the noise free recordings. Statistical multivariate analysis was performed and accurate characterization of the restoration techniques used was achieved. In order to correlate perceptual results with acoustic characteristics of the different stimuli, a time-frequency analysis was performed. These results permitted to relate the esthetic judgement categories used by listeners with the main features of the restoration techniques used.

KEYWORDS: Audio Restoration, Digital Signal Processing, Perceptual Tests.

INTRODUCTION

The cultural heritage of musical recordings is slowly fading away since the physical media, where the audio signal information is stored, are exposed to deterioration.

The introduction of digital audio carried out remarkable advantages in the field of music preservation and restoration. In fact, the copies digital-to-digital do not suffer of the analog-to-digital re-recordings problems: noise addition, difficulty to fix the 0dB level position, equalization curves choice to set the tape (or vinyl) player. Moreover, in digital domain, it is practicable to carry out the necessary periodic re-recordings in an automatic way. It is also feasible to use algorithms finalized to noise-removal, more sophisticated than analog systems like Dolby B, Dolby C or dbx.

The audio signal is subject to two kinds of technical alterations [9]: unintentional and intentional. The first group is caused by the imperfection of the recording techniques (resulting in various distortions) and misalignment of the recording equipment (resulting in wrong speed or azimuth errors). The group of intentional alterations concern recording equalization, noise reduction systems, and so on.

Furthermore, dirt as well as physical and chemical deterioration have various influences on carrier, ranging from less obtrusive clicks on mechanical carriers to shorter or longer losses of signal, and linear or non-linear distortions. Causes of degrade can have local influence (as scratches and dust) or global influence (as white noise, distortions, wows and flutters). In this scenario, it is getting more and more important to devise different techniques aimed at digital audio restoration. In this sense, the aims of this work are:

- to define a subjective evaluation method of the restoration quality of corrupted audio signal and to apply this method to the comparison of different digital restoration techniques;

- to observe the different listener's expectation, i.e., to understand if the listener (musically-trained or musically-untrained) pays attention to the reduction of noise level, and/or to the artifacts (distortions or lowpass-effects) introduced in the music signal by the restoration process.

In the paper we first shortly overview the most used audio restoration methods. Then we present the setup of the comparison carried out and a time-frequency analysis of the restored stimuli to characterize the algorithms used. A final part is dedicated to relate the perceptive analysis with the results of the acoustic analysis.

AUDIO RESTORATION METHODS: AN OVERVIEW

Time-domain algorithms:

The methods based on Extended Kalman Filter (EKF) theory are based on AR modelling of the data [8]: EKF, in principle, solves the problems of broadband noise filtering, signal parameters tracking and impulsive noise

removal [7]. At medium/high signal-to-noise ratios (SNR), the performance of such filter is superior to that of other standard methods like Spectral Attenuation, nevertheless its plain use does not guarantee the best results. One reason for this is that the nonstationarity of audio signals leads to errors in parameter and tracking noise filtering, especially during fast transients. In order to achieve maximum performance from the EKF, it is essential to optimize its implementation [1].

Frequency-domain algorithms:

Adaptations of the spectrally-based techniques applied in speech processing have received most attention. They are based upon spectral weighting, in which individual spectral components are weighted according to expected noise and signal components. Such techniques can be viewed as finite block-size approximations to frequency domain Wiener filtering. As a result of these approximations (necessary to follow the time-varying nature of the useful signal) undesirable distortions can occur, the most notable being known as "musical noise" in which statistical fluctuations in the frequency components of noise lead to random tonal artifacts in the processed signal. Various techniques have been applied to mask or eliminate these distortions. We adopted the Ephraim and Malah suppression rule [5].

Algorithms based on psychoacoustic models:

This task requires to transform the audio signal from an "outer" to an "inner" representation, that is to resort to a representation that takes into account how the human ear perceives the sound. The combination of the psychoacoustic model and Frequency-domain algorithms, permits to define a promising restoration methodology [4].

EXPERIMENT

To single out the best computational methodologies for audio restoration, a perceptual test was made, using commercial software and our implementations of the methods presented in the last section.

Four noise free recordings were intentionally corrupted with different real (i.e., not synthetic) scratches and broadband noises extracted from tapes and vinyl disks. The noise free recordings were selected to cover a wide range of musical genre (Jazz, Western Classical, Pop) and different recording age (from 1940 to 1970); the stimuli were played with different instruments (piano solo, symphonic orchestra, jazz orchestra, voice) in order to analyze the deterioration persistence related to the different physical characteristics of audio signal (bandwidth, transients, timbre).

Finally, the noisy stimuli were restored using commercial software (Dart Pro, CoolEdit Pro, Wavelab) and proprietary software based on Extended Kalman Filter [1], [7] and on a psychoacoustic model [2], [4]. The parameters used to control the different systems were subjectively setup to obtain the best tradeoff between noise-removal and music-signal-preservation.

Twelve subjects were asked to score (from 0 to 5) the restoration quality compared to the original 'noise free' recordings quality. Two subject groups were selected: the first one constituted by researchers (of the Centro di Sonologia Computazionale, University of Padova) and the second one by skilled students. All the subjects were musically trained. The method used to define the evaluation criteria is an extension of the perceptual experiments

described in the International Communication Union Protocols [ITU-R BS.1116 and ITU-R BS.708] on digital signal quality.

This protocol is based on "double-blind triple-stimulus with hidden reference" method, which is stable and permits accurate detection of small impairments. In the preferred and most sensitive form of this method, one subject at a time is involved. The known reference ("noise free" stimulus) is always available as stimulus "A". The hidden reference ("noise free" stimulus, a copy of stimulus "A") and the object (restored stimulus) are simultaneously available but are "randomly" assigned to "B" and "C". The subject is asked to assess the impairments on "B" compared to "A", and "C" compared to "A", according to a continuous five-grade impairment scale. As soon as the subject has completed the grading of a trial, it should be possible to proceed directly on to the next trial. The excerpts "A", "B" and "C" may be repeated in any order until the subject makes an assessment. In this way the test procedure is self pacing. The grading scale shall be treated as continuous with "anchors" derived from the ITU-R five-grade impairment scale given in Recommendation ITU-R BS.562 (Imperceptible - Perceptible, but not annoying - Slightly annoying - Annoying - Very annoying).

Different factor analyses were performed on subjects answers to characterize the behavior of the audio restoration algorithms. An acoustic (in time-frequency domain) analysis was carried out on the restored signal to relate the results of factors analysis with the physical characteristics of sound. In this way it was possible to highlight the acoustic parameters that are more

important from a perceptual point of view.

TIME-FREQUENCY ANALYSIS

Fig. 1 shows the spectrogram of an excerpt (4 s long) of “Johnny comes lately” (rec. 26 June 1942, Duke Ellington Orchestra, CD Bluebird 5659-RB). Fig. 2 shows the spectrogram corresponding to the addition of broadband noise extracted from a tape. Fig. 3 and fig. 4 give the spectrograms of the restored stimuli using, respectively, Cool Edit PRO and DART PRO. It can be noticed that in figure 4 (restored by DART PRO) there is an evident low-pass effect (around to 6.5 kHz), which gives an audible distortion to the signal. On the contrary, the spectrograms show that Cool Edit PRO (fig. 3) does not remove all the broadband noise, unlike DART (see fig. 4). The measurement of SNR shows that DART improves the value of 10dB. In this case, from perceptual test (see next section) emerges a different listener’s expectation, i.e., a cluster of listeners gives importance to the reduction of noise level and seems to prefer low broadband noise (so they prefer DART), and another cluster pays attention to the artifacts (lowpass-effects, in this case) introduced in the music signal by the restoration process (choosing Cool Edit PRO).

The behavior of the other considered methods is intermediate between the one of Cool Edit PRO and DART PRO and will be confirmed by the results of the following section.

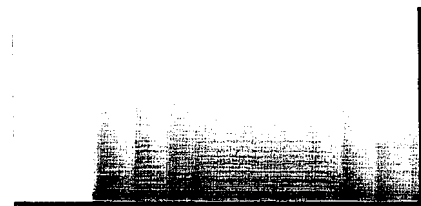


Fig. 1: spectrogram of “Johnny comes lately”, from 0 to 4 s (“noise free”). The y-axis (frequency) is scaled from 0 to 15kHz. The x-axis represents the time in s

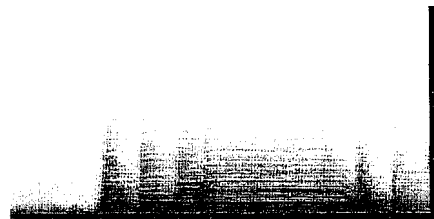


Fig 2: spectrogram of the first 4 seconds of “Johnny comes lately” with broadband noise (“tape”) added. The y-axis is scaled from 0 to 15kHz

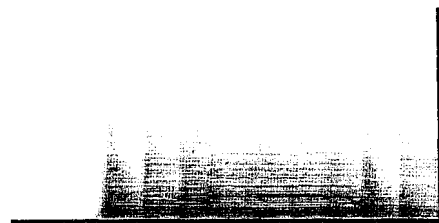


Fig 3: spectrogram of “Johnny comes lately” restored from broadband noise with Cool Edit Pro). The y-axis is scaled from 0 to 15kHz

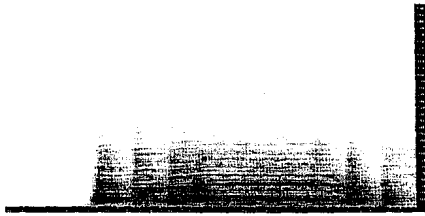


Fig 4: spectrogram of “Johnny comes lately” restored from broadband noise with DART Pro. The y-axis is scaled from 0 to 15kHz

The impulsive disturbances (click and scratches) are totally removed by all systems used. The signal masked by the click is reconstructed by interpolating the missing samples according to different strategies, introducing an evident signal deterioration.

PERCEPTUAL ANALYSIS

A factor analysis was carried out to homogeneously group the subjects, using Principal Component Analysis; the subjects were considered the observable variables. This analysis shows two principal factors that explain, respectively, 55.8% and 9.1% of the global variance. Factor 1 appears to be related to the subjects attention or cooperation during the test. Factor 2 is related to the listeners experience (or competence). Plotting, as in Fig. 5, the factor loadings in a two dimensional space helps to analyze the homogeneity among the subjects. The figure clearly shows that there are two clusters of subjects: the first one containing the researchers and the second one including the students. Note that the first factor has a high value for both groups.

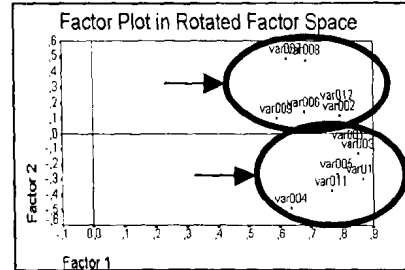


Fig. 5: factor loading plot, taking the subjects as variables. Factor 1 (55.8% of total variance) is related to the subjects attention during the experiments (inattention vs. attention). Factor 2 (9.1%) is related to listeners experience (incompetence vs. competence)

In order to observe the persistence of different noises, another analysis on the data was made transposing the answer matrix and carrying out a factor analysis taking the added disturbances as variables. This time, two main factors emerge: the first one explains 46.7% of total variance and the second one 20.1%. The acoustic analysis relates the first factor with the noise reduction carried out by the restoration methods, and the second factor with signal quality loss. Fig. 6 gives the corresponding factor plot for the different noise types considered: scratch, click and broadband noise. The more persistent deterioration is the impulsive noise (scratches and clicks), which has both a high value of signal quality loss and a low value of noise reduction. The broadband noises used were, respectively, disk noise (D-noise) and tape noise (T-noise) with Signal Noise Ratio 25dB (noise1) and 35dB (noise2).

Finally, the algorithms were considered the observable variables of factor analysis.

Four factors were deduced and Tab. 1 gives the factor loadings obtained by factor analysis (for factors 1, 2, 3 e 4).

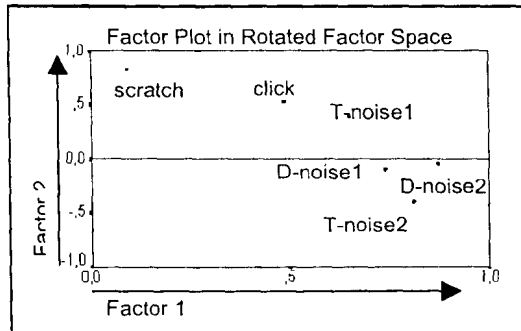


Fig. 6: factor loading plot, taking the different disturbances as variables. Factor 1 (46.7.8% of total variance) is related to the deterioration attenuation. Factor 2 (20.1%) is related to the signal quality loss

Factor 1 (38.5% of total variance) and factor 2 (17.2%) are probably related to click and scratch removal. In fact all the algorithms have similar behavior: removal of impulsive noise that is accompanied by a (perceptually relevant) signal distortion. Factor 3 (14.7%) is related to the distortion caused by broadband noise removal (high vs. low distortion). In fact, Cool Edit has the maximum factor loading (0.85, see tab. 1) and it is the software that introduces the minimum distortion to the signal spectrum. An old (proprietary) algorithm based on Wiener filter (W1*) has the minimum factor loading (-0.28, see tab.1): it is the result of a old experiment finalized to the noise attenuation, without respect to the signal distortion.

Factor 4 (11.1%) is related to broadband noise attenuation (low vs. high attenuation). In this factor DART has a maximum factor loading (0.89, see tab.1): in fact, it is the software that obtains the best SNR.

Factor	Commercial Algorithms			Proprietary Algorithms				
	C	D	WL	P1	P2	W1	W2	K
1	.01	.02	-.01	.0	-.01	.01	.02	.04
2	.03	.02	-.01	.0	-.02	.0	.01	.03
3	.85	.12	.28	.04	-.12	-.28	.45	.33
4	.05	.89	-.25	.13	.28	.29	.18	-.05

Tab. 1: factor loadings for the algorithms used in the perceptual test. Factor 1 (38.5% of total variance) and factor 2 (17.2%) are related to the click and scratch removal; factor 3 (14.7%) is related to the distortion caused by broadband noise removal (high vs. low distortion); factor 4 (11.1%) is related to broadband noise attenuation (low vs. high attenuation). C = Cool Edit Pro, D = Dart Pro, WL = Wavelab, Wn = different implementations of Wiener filtering, Pn = different implementation of a psychoacoustic model, K = Kalman filtering.

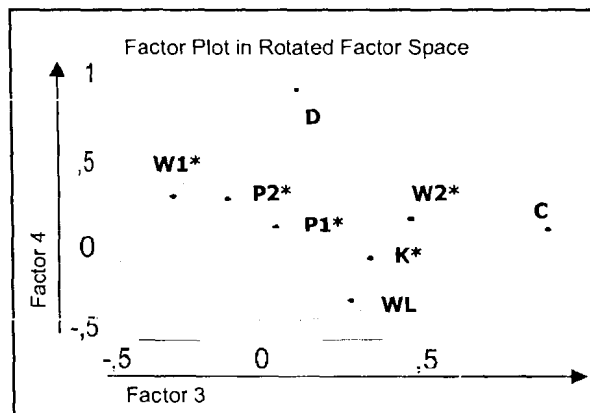


Fig. 7: factor loading plot, taking the different algorithms as variables. Factor 3 (14.7%) is related to the distortion caused by broadband noise removal (high vs. low distortion); factor 4 (11.1%) is related to broadband noise attenuation (low vs. high attenuation). C=Cool Edit Pro, D=Dart Pro, WL=Wavelab, Wn=different implementations of Wiener filtering, Pn=different implementation of a psychoacoustic model, K=Kalman filtering. (*=proprietary algorithms)

Fig. 7 plots the two dimensional space generated by factors 3 and 4. The variables (algorithms) present high correlation index in the first two factors that, consequently, are not plotted.

It can be noticed that, in relation to the distortion caused by broadband noise removal, the best algorithms are those based on Short Time Spectral Attenuation (C and W2*) [6] and on Kalman filter (K*) [1]. In relation to broadband noise attenuation, Dart Professional (D), Wiener filter (W1*), and the algorithm based on psychoacoustic model (P*) ([2], [4]) present a better behavior that is in good accordance with the frequency analysis presented in the last section.

CONCLUSION

The aim of this work was to introduce an evaluation method of the restoration quality of noisy audio signal and to apply it to the comparison of different digital restoration techniques.

A comparison between proprietary algorithms and some commercial products was made on the basis of a perceptual test. The method used to define the evaluation criteria is an extension of perceptual experiments described in Int. Com. Union Prot. [ITU-R BS.1116 and ITU-R BS.708] on digital signal quality. Statistical multivariate analysis was performed and accurate evaluation of the restoration techniques used was achieved. These results permitted to relate the esthetic judgement categories used by listeners with the main features of the restoration techniques.

A factor analysis taking the "subjects" as variables was performed to evaluate if the listening ability was uniformly distributed across all the subjects.

Another factor analysis was carried out using the "corrupted musical excerpts" as variables, in order to weight the influence of the disturbances on the signal. From this analysis emerges that the impulsive noise is the disturb most annoying for listeners: moreover, its removal introduces an audible signal quality loss (see fig. 6).

A further factor analysis, using the "restored musical excerpts" as variables, led to classify the effectiveness of the algorithms used with respect to the restoration quality. Each algorithm was classified focusing on the behavior with respect to the different classes of noises and different audio categories.

From the analyses it emerged that it does not exist an optimum audio restoration algorithm, able to remove, in a satisfactory way, all the different disturbance typologies without introducing some quality loss into the audio signal. In particular, some algorithms (e.g., D, W1*, W2* and P2* of fig. 7) are able to completely remove disturbances, but at the expense of signal quality loss (as exemplified in fig. 4). At contrary, other algorithms (e.g., C and K* of fig. 7) are more "respectful" of the useful signal, but do not greatly attenuate noise (e.g. compare fig. 3 and fig. 4).

ACKNOWLEDGMENTS

This work was supported by CNR – "Progetto Finalizzato Beni Culturali".

REFERENCES

1. Bari A., Canazza S., De Poli G., Mian G. A., (1999). "Some key points on restoration of audio documents by the Extended Kalman Filter". Proc. of Diderot Forum, Vienna, 2-4 December 1999, 37-47.

2. Beerends J. G and J. A. Stemerink, (1992). "A Percentual Audio Quality Measure Based on Psychoacoustic Sound Representation". *J. Audio Eng. Soc.*, 40(12), pp. 963-978.
3. Boll S. F. and A.V. Oppenheim, (1979). "Suppression of acoustic noise in speech using spectral subtraction". *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-27(2), April 1979.
4. Canazza S., De Poli G., Maesano S., Mian G. A., (1999). "On the performance of a noise reduction technique based on a psychoacoustic model for the restoration of old audio recordings". *Proc. of Diderot Forum, Vienna*, 2-4 December 1999, pp. 29-35.
5. Ephraim Y. and Malah D., (1984). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". *IEEE Trans. Acoustics, Speech and Signal Processing*, 21(6) pp. 1109-1121.
6. Godsill S., Rayner P., Cappé O. (1998). "Digital audio restoration". In *Applications of digital signal processing to audio and acoustics*. Kahrs - Karlheinz Brandenburg (ed.). Kluwer Academic Publishers.
7. Niedzwiecki M. and Cisowski K. (1996). "Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals". *IEEE Trans. Signal Processing*, 44(3), pp. 287-300.
8. Paliwal K. K. and A. Basu (1987). "A speech enhancement method based on Kalman filtering". *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 177-180.
9. Schueller D. (1991). "The ethics of preservation, restoration and re-issues of historical sound". *J. Audio Eng. Soc.*, 39(12), pp. 1014-1016.

About the authors

ABOUT THE AUTHORS

Sergio Canazza is researcher at the "Centro di Sonologia Computazionale" (CSC), University of Padova. In 2000 he's been teaching in Master "Transfer of Multimedia Technology in Small and Medium Enterprise system in the field of Cultural Heritage", at the Federico II University of Napoli. Since 1998 to 2001 he's been participating in the research for the CNR "Progetto Finalizzato Beni Culturali" and the "Restauro filologico dei nastri magnetici delle opere di Bruno Maderna" supported by MURST.

E-mail: canazza@dei.unipd.it

Giovanni De Poli is Professor in Computer Science at Dept. of Electronics and Informatics, University of Padova. He is director of Centre of Computational Sonology (CSC). Since 1998 to 2001 he's been participating in the research for the CNR "Progetto Finalizzato Beni Culturali".

E-mail: depoli@dei.unipd.it

Gian Antonio Mian is Full Professor in Digital Signal Processing at Dept. of Electronics and Informatics, University of Padova. Since 1998 to 2001 he's been director of the CNR contract "Progetto Finalizzato Beni Culturali".

E-mail: mian@dei.unipd.it