



HAL
open science

ON THE USE OF VARIABLE LENGTH WINDOWS IN THE RESTORATION OF AUDIO RECORDINGS

O. Cappé

► **To cite this version:**

O. Cappé. ON THE USE OF VARIABLE LENGTH WINDOWS IN THE RESTORATION OF AUDIO RECORDINGS. Journal de Physique IV Proceedings, 1992, 02 (C1), pp.C1-101-C1-104. 10.1051/jp4:1992118 . jpa-00251141

HAL Id: jpa-00251141

<https://hal.science/jpa-00251141>

Submitted on 4 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE USE OF VARIABLE LENGTH WINDOWS IN THE RESTORATION OF AUDIO RECORDINGS

O. CAPPÉ

*Département Signal, TELECOM Paris, 46 rue Barrault, F-75634 Paris cedex 13, France***Abstract**

The application of spectral subtraction methods to the restoration of audio recordings degraded by background noise generally yields satisfactory results. However, these methods introduce some distortions that can be quite audible in the case of highly degraded recordings. In the first part of the paper it is shown that some of these distortions are related to the use of a noisy phase spectrum. A solution to the problem of transient signal smearing is then presented. This new procedure makes use of spectral subtraction on variable length windows in conjunction with a method for detecting rapid energy variations.

1 Introduction

In the restoration of audio recordings degraded by background noise there is generally no available reference source for the degrading noise. Moreover, due to the variety of musical signals to be restored, it is difficult to make any assumptions on the signal. We will consequently focus on spectral subtraction, where it is only supposed that the degradation can be modeled as uncorrelated additive stationary noise. "Spectral subtraction" stands for the group of methods based on short-time Fourier transform (STFT) analysis/synthesis performing a modification of the short-time amplitude spectrum. This type of methods has been previously applied to the restoration of degraded musical recordings [1]. But one of the limitations of spectral subtraction is that it creates perceptual distortions in the restored signal [2]. In the first part of the paper it is shown that some of these distortions are due to the STFT processing itself and not to a particular noise reducing strategy. Then some modifications of the basic procedure are suggested in order to improve the quality of the restoration.

2 Perceptual influence of the short-time phase**2.1 Spectral subtraction procedures**

The purpose of spectral subtraction methods is to estimate the short-term spectral amplitude corresponding to the noise-free signal [3] [4] [5]. The attenuation curve applied to the noisy short-term spectrum in order to obtain the spectral amplitude estimate is known as the spectral subtraction gain. The estimated short-term magnitude is associated with the noise corrupted short-term phase in order to reconstruct an enhanced signal. It is assumed that the STFT parameters are such that this synthesis procedure is valid [1]. However, a preliminary study [6] highlighted the fact that the presence of additive noise on the short-time phase spectrum creates audible distortions in the restored signal.

2.2 Distorsion for periodic sounds

We consider the case where the signal to be restored consists in a periodic sound degraded by additive non-correlated noise. The distortion due to the use of the noisy phase spectrum is then efficiently described as a noise-modulation of each of the sound's harmonic components. The modulation characteristics (power and frequency span of the modulating noise) depend on the original additive noise statistics but also on the window duration used for the STFT [6]. Thus for commonly used STFT settings (overlap-add procedures with a window overlap of 50% to 75%), an increase in the window duration T has two effects:

- First, the modulating noise power is reduced by a factor $(1/T)$, since the level of the incoherent components (the noisy parts of the spectrum) are lowered with respect to the tonal components.
- The frequency span of the modulating noise is limited by approximately the same factor $(1/T)$.

The audibility of the modulation effect depends on the respective frequencies of each component of the sound, but also on the listening conditions (the SPL level in particular) [7]. The results can be summarised as follows:

- For window durations of 20 ms and less,

Some roughness appears in the restored signal even when the original signal-to-noise ratio is relatively high (more than 10 dB). This effect is stronger for low frequency components because the associated critical bandwidth is narrower [7]. Moreover, the distortion is emphasized if there is some amount of vibrato in the original sound.

- For window durations between 20 ms and 80 ms,

An increase in the window duration is clearly perceived as an improvement. The effect of such an increase is to reduce the modulation effect for low amplitude components (local signal-to-noise ratio of less than 10 dB). The improvement is particularly clear when slowly fading sounds are present, because in this case the signal-to-noise ratio decreases, making the distortion more and more audible.

2.3 Temporal smearing of transient sounds

On the other hand, when a rapid transient signal is present (in percussive sounds for example), the use of noisy phase information induces a temporal smearing in the STFT windows where the transient signal occurs [2]. This effect, referred to as "delocalization", can be seen in the spectrograms of figure 1. The processed signal exhibits only a low level residual noise, but the stretching of the transient is clearly visible.

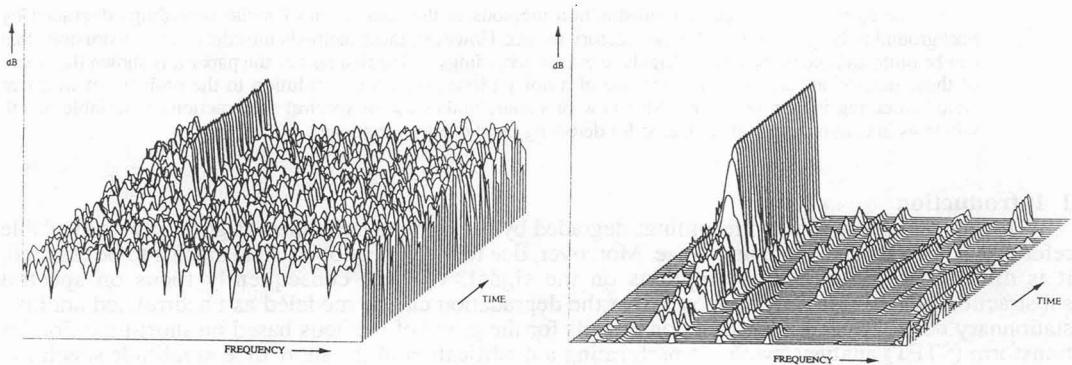


Figure 1: High (2 ms) temporal resolution spectrogramms (1)-50 ms of the abrupt onset of a pure tone degraded by white noise, and (2)-same signal after restoration by spectral subtraction (duration of the window used for restoration, $L = 30$ ms).

This blurring effect alters the timbre of transient sounds and is sometimes perceived as the result of artificial (and non causal !) reverberation. The delocalization of a transient event is all the more severe as the transient's energy gets lower compared to the power of the noise. However, the smearing of a given transient signal is limited in time to the window duration.

3 Variable length processing windows

3.1 Motivations

The aim of the present section is to elaborate an analysis/synthesis setting based on STFT that allows an efficient noise reduction with the lower possible distortion. This means that we need simultaneously long analysis windows in order to prevent the noise reduction procedure to introduce a modulation effect on stationary sounds, and high temporal resolution to avoid the delocalization of transient sounds.

To meet this two contradictory requirements, we propose to use variable length windows for the STFT. The point is that the choice of a window-duration for a given portion of the recording depends on the recorded signal itself. We must point out that analog conclusions concerning the effects of window-duration led to a completely different implementation based on a prior subband analysis in [2].

3.2 Description of the processing

The procedure is the following:

- in general, a long window-duration is used (typically 60 to 80 ms).
- when transient variations of the signal are detected, a shorter window (20 ms) is selected.

The short window-duration is used only in the vicinity of the detected event (during approximately 80 ms). For both window-durations, the noise reduction is achieved through spectral subtraction by use of the algorithm proposed in [4]. The continuity between these two different window-durations is achieved by the respect of the overlapp-add conditions relative to the long window-duration. An example of the switch to a shorter window-duration can be seen on figure 2.

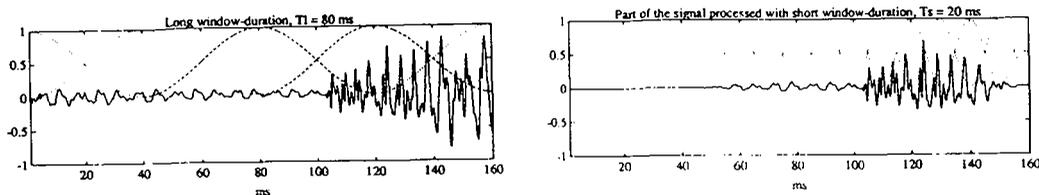


Figure 2: Switch from long window-duration to short window-duration when a transient signal is detected.

Except for some border effects, it is clear from figure 2 that this procedure is equivalent to processing the degraded signal twice, once with the short window-duration and once with the long one. The two restored signals are added afterwards, with proper weighing, around the position of the detected transient events. This computationally expensive off-line procedure can easily be used for simulations.

3.3 "musical noise"

Some special care must be taken when the short window-duration is used: It is a well known fact that spectral subtraction leads to a colored residual noise known as musical noise. Some procedures reduce the disturbance due to this noise [3] [4], but they are less efficient when a transient signal is present because they rely on the redundancy from one analysis window to the other. Moreover, the level of this musical noise is proportionally higher when a short window-duration is used for the STFT, because the Fourier transform hardly separates tonal components from noise in this case. Consequently, a straightforward application of the proposed procedure may lead to the appearance of small bursts of musical noise when musical transients occur.

To avoid this effect, it is preferable to use a low resolution spectral subtraction when the short window-duration is used. This implies that the power spectral density of the noisy signal is estimated by use of the averaged periodogram. The analysis window is typically splitted into 2 to 4 sub-windows which are zero-padded, Fourier transformed and averaged. A spectral subtraction gain with low frequency-resolution is obtained, but the variance due to the noise is greatly reduced. This gain curve is then applied to the high frequency-resolution short-time spectrum obtained over the whole analysis window as in standard modification of the STFT. The noise reduction is less efficient (the noise is still present in some frequency areas), but the musical noise is greatly reduced. For our application, the loss of accuracy of the noise reduction is not a problem because the short window duration is used only when transients are detected in the signal.

4 Detection of events

4.1 Nature of the events to be detected

The procedure described previously requires prior detection and localization of rapid variations in the musical signal. On the basis of our experience, the delocalization of musical transients is mainly perceptible for sounds with very brief starting transients. Therefore, only the events corresponding to rapid variations of the signal's energy need to be detected.

4.2 Detection test

The adopted procedure consists in:

- First, computing the signal's energy level on a dB scale, with temporal window-durations of approximately 80 ms. The window increment should be sufficiently small (typically 2 to 5 ms) in order to allow a good localization of the events.

- The variation of the dB-energy level (called LEV) is then used as the change indicating signal for the detection. The CUSUM test is used for monitoring the LEV signal. The purpose of the CUSUM test is to detect (and localize) jumps in the mean of a stochastic variable [8]. Here, the mean value of the LEV signal is estimated recursively using a fading memory (the forgetting factor is usually chosen around 0.998). The introduction of a fading memory in the energy variation predictor allows to follow

slow energy drifts while detecting only the rapid energy variations. Figure 3 displays the processed signal and the corresponding mean LEV value. The vertical jumps indicate the detection times.

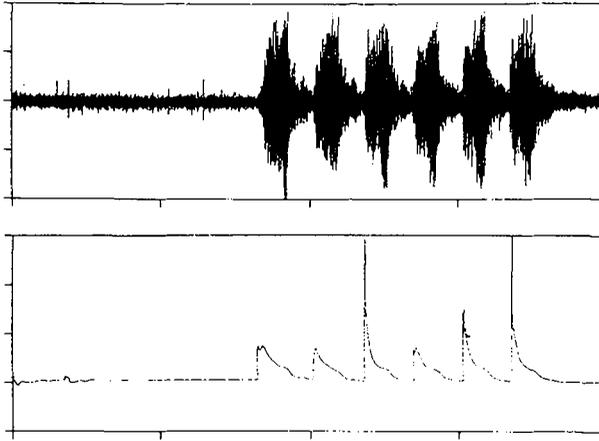


Figure 3: Detection of localization of rapid energy variations (total time displayed 3,2 s)

5 Results

The test for detection of rapid energy variations was found to be very robust, and it allows a good localization of the detected events (typically the precision of temporal localization is a few times the value of the window increment). Since the criterion of global energy variation may not be sufficient for complex musical contexts, we are currently evaluating a generalization of this detection procedure in subbands. The complete procedure was applied on different degraded recordings. Listening tests indicate that the transients restitution is significantly improved in the case of strongly degraded recordings (signal-to-noise ratios below 15 dB). The improvement is particularly clear in the case of percussive sounds, where the best restitution of the starting transient gives more brightness to the restored sound. However, it must be noted that for such highly degraded recordings, the quality of the restoration is not yet fully satisfactory, particularly in terms of timbre homogeneity between the different intensity nuances of the musical sound.

References

- [1] R.Bourdier, "Analyse temps / fréquence, filtrage et synthèse numériques de signaux de parole. Application au filtrage, à la réduction de bruit, et à la restauration d'enregistrements anciens", Thèse, Université du Maine, Le Mans, 1988.
- [2] J.C.Valière, S.Montessor, J.F.Allard, "Présentation d'une méthode de suppression des bruits de surface sur les anciens enregistrements de musique", 1er Congrès Français d'Acoustique, Colloque de Physique C2, supplément au n°2, tome 51, p 761 - 764, Février 1990.
- [3] S.F.Böll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. ASSP, vol 27, n° 2, p 113 - 120, 1979.
- [4] Y.Ephraim, D.Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator", IEEE Trans. ASSP, vol 32, n° 6, p 1109 - 1121, 1984.
- [5] P.Vary, "Noise suppression by spectral magnitude estimation - Mechanism and theoretical limits -", Signal processing, vol 8, n° 4, p 387 - 400, 1985.
- [6] O.Cappe, A.Chaigne, "Perceptual effects of noise disturbances on phase spectrum in STFT analysis/synthesis procedures. Application to restoration processes.", 1991 IEEE ASSP Workshop on applications of signal processing to audio and acoustics, New York, October 20-23, 1991.
- [7] E.Zwicker, R.Feldtkeller, "Psychoacoustique, L'oreille récepteur d'information", Masson, 1981.
- [8] M.Basseville, "Detecting changes in signals and systems - A survey", Automatica, vol 24, n° 3, p 309 - 326, 1988.