# FREQUENCY-DOMAIN PROCESSORS FOR EFFICIENT REMOVAL OF NOISE AND UNWANTED AUDIO EVENTS

**CHRISTOPH M. MUSIALIK and ULRICH HATJE**

*Algorithmix GmbH*
*Klettgaustrasse 21, 79761 Waldshut-Tiengen, Germany*
cmm@algorithmix.com hatje@algorithmix.com

Filtering heavily polluted audio signals with low-cut, high-cut or notch filters helps only if the spectrum of disturbances does not overlap the desired signal. Unfortunately, this is rarely the case in real situations. Therefore using frequency-domain methods is the only way to achieve dramatic improvements when cleaning up noisy signals. Spectral denoising works well in the case of broadband, smoothly changing noise, however, the situation become hopeless when the noise character changes abruptly and the S/N ratio is around zero decibels. We present a unique tool, reNOVAtor, working in 3D domain (time, frequency, amplitude) allowing efficient removal of tenacious disturbances. An intuitive man-machine interface allows localization, identification, and very precise removal of unwanted audio events. Special functions like automatic detection of clicks, tones, and harmonics significantly accelerate the workflow. A special difference function allows monitoring of the signal portion being removed.

## 1. TYPES OF AUDIO DISTURBANCES

Noise and discrete disturbances are among the most common problems in the daily business of an audio engineer, who needs to deal with the whole range of unwanted signal disturbances, especially when restoring old recordings including ancient wax cylinders, 78 rpm (shellac) and vinyl records, different kinds of tapes (including cassettes) and optical film tracks. Historically, we divided them into three groups: clicks and crackles, broadband noise and wow & flutter [1, 2].

Clicks are short impulsive disturbances typical for old records. They can also be caused by hard switching of audio sources, cross-talk from digital cables and thyristor controlled light installations. Crackles are associated with small clicks which are typical for shellac and vinyl records. While typical clicks can be successfully removed with different products on the market, crackles are not that easy to detect and remove, because they often cannot be uniquely distinguished from high-slope low-level original signal details.

Wow & flutter are disturbances that cause pitch changes. They are related to mechanical problems like irregularities in rotating elements or material deformations (records, tapes, wax cylinders). These kinds of distortion require very specific restoration methods which are not related to the technologies described in this paper.

Normally we associate noise with tape hiss or electronic hiss typical for any kind of (pre)amplifiers. In such cases the specification of noise is quite easy. Also air conditioning equipment noise or constant buzz and hum are still easy to describe, because we have an almost stationary kind of noise. The situation becomes much more complicated if the noise level and its character varies, or if only a part of noise is not wanted, but the other part belongs to the ambience we want to recover.

The highest degree of difficulty in audio restoration occurs when we have to deal with disturbances which cannot be strictly classified to one of the two historical groups: impulsive disturbances (clicks) or broadband (stationary) noise. In practical situations we are confronted with very broad gray zone: drop-outs, thumbs (low-frequency ‘clicks’), non-linear signal distortions (clipping) and finally longer unwanted complex audio events. In this paper we present tools which allow coping with such ‘non-standard’ disturbances.

## 2. MUSIC vs. FORENSIC AUDIO RESTORATION

Audio restoration was originally developed to clean up historical music recordings. Based on digital technology some successful products have been created for restoring old recordings. Most of them can remove unwanted clicks and broadband noise quite well, but it is not everything that professional re-mastering engineers expect from such systems today. It is not only important to remove disturbances, but also to keep original material audibly untouched and free from newly added artifacts. The reality is, however, that most systems produce artifacts and the original signal is in some way affected. The most typical problem is the lack of high frequencies and ambience in the restored signal.

The situation becomes even more challenging if we master live recordings done with state-of-the-art equipment, thus with the best possible quality today. Let's imagine a unique classical concert. After checking the recorded material in your studio before production and mastering you realize that the recording was successful in general, except for a few annoying disturbances during some quiet passages: somebody's cough, a squeaky chair, the horn of a passing truck, a bell from the neighboring clock tower. In addition, despite the exceptional artistic interpretation there were a few significant errors: a loud scratch in the part of violin soloist and one too early tone in the brass section.

All this makes your recording unacceptable and, of course, the concert cannot be repeated. As an experienced tonemeister, you know very well that all traditional techniques and tricks fail when you try to remove the disturbances mentioned above. In such a situation any kind of traditional equalization or sophisticated editing method is usually time consuming and causes discontinuities or at least audible changes in level and timbre of the desired signal and ambience.

In forensic audio the audio quality is usually not as important as in cleaning of artistic performances. There we deal typically with speech dialogs recorded under very poor conditions, often with hidden microphones or very far away from the person we want to record. Frequently the signal-to-noise ratio is around zero or even below. The most important point is to increase speech intelligibility in order to understand spoken words. It is often not important if the restored material includes newly added artifacts provided the understandability has been significantly increased.

However, we do have situations when the sound quality becomes important in forensic audio. Material cleaning before speaker identification or isolation of important audio events like gunshots, noise of a particular car, or specific details in the acoustical ambience.

Summarizing, the sound restoration of artistic performances is optimized to removal of unwanted noises without touching the original sound and its ambience integrity and to avoid introducing new artifacts during the restoring process. In the case of forensic audio, sound quality and artifacts (especially if they are not sonically annoying, which is the case after even exaggerated denoising) are not as important. The difficulty for the algorithms results here from the significantly poorer recording quality and SNR around or even below zero.

## 3. NOISE REMOVAL

We mentioned different kinds of noises. The easiest to deal with is broadband noise that can be defined as spread over the whole signal spectrum. If its energy in analyzed frequency bands is constant over time we can speak about stationary noise such as tape hiss and amplifier noise. They are perceived as having more energy in high frequencies. The most primitive way to denoise is to use a low-pass filter which reduces the noise energy above its cut-off frequency. Of course, if the desired signal itself contains high frequency components, they are also lowered. Better performance is provided by a dynamic low-pass filter that cuts the noise only if the original signal is not present. If the desired signal is loud enough it masks the noise automatically. Unfortunately, even properly designed dynamic filters have internal inertness (need some time to react) and thus smear fast signal transients.

Another tool to suppress perceptible noise is the noise gate. It mutes the output if the input signal drops below a certain threshold. It has only two states--therefore the proper transient treatment is difficult to implement here. Usually, better results can be achieved using an expander that works similarly to the noise gate, but can apply progressive gain reduction dependent on the threshold and ratio setup instead of fast and complete muting of the output signal. A further improvement is provided by a multiband expander that acts on the signal divided into some frequency bands, but here the problem is also that not only noise in every channel, but the desired signal too, is removed by the same reduction factor (set up by the user) independent of the signal level--if we reduce the noise by 35% we also reduce the signal by 35%. The reason is that even quite sophisticated expanders are unable to distinguish between noise and desired signal and thus they also remove parts of the audio material that must be recovered. Ideally, we would wish to subtract the same amount of noise and keep the desired signal untouched, independent of the input signal level. This became more realistic after the introduction of DSP technology.

In the last twenty years, various DSP noise reduction algorithms have been developed, originally for speech enhancement in transmission channels. They differ in noise removal efficiency, computational power and amount of introduced artifacts. To date, the best compromise among all these factors can be achieved by using the so-called short-time spectral subtraction [4, 10, 2]. Enhanced by some techniques for artifact suppression [4, 7 - 18], this method can be successfully applied for a wide range of applications: from the high-quality denoising of valuable music treasures to the treatment of forensic material recorded at a very poor signal-to-noise ratio and/or in a reverberant environment.

We assume that the audio signal to be restored is the sum of the original audio signal and an uncorrelated additive noise signal. The only observable signal is the degraded signal. Comparing to the complementary noise reduction methods when the original signal is pre-processed or coded before degradation (like in Dolby system or other companders) we classify this method as single-ended noise reduction systems.

The idea of short-time spectral attenuation is shown in Figure 1. The corrupted input signal is first analyzed by short-time FFT (or in general any kind of multi-rate filter). Then every bin is attenuated by a certain factor coming out of the decision system block. Finally, the modified short-time spectral representation is transformed back with inverse FFT to obtain the restored signal. The amount of the applied attenuation is calculated from the estimated power of noise and suppression rules [8, 9] related to the character of noise and a selection of preferences (e.g., less denoising and less artifacts, better speech quality but more artifacts).

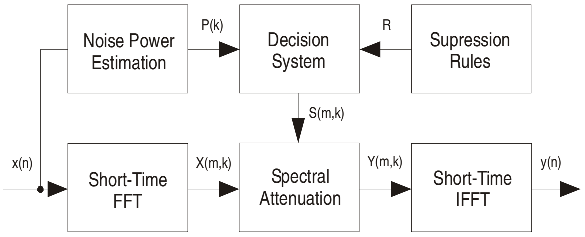


Figure 1: Denoiser using spectral attenuation.

Often the short-time noise suppression is considered as delivering moderate results, because of either too little suppression or too many artifacts added after processing. This is true if we implement this method in the school-book form. There are some significant improvements possible if we pay attention to the spectrum estimation methods [8, 10], suppression rules [4, 9] and the way the time-varying attenuation is implemented. Here psychoacoustical knowledge can help a lot [15, 16].

Short-time Fourier transform allows almost perfect reconstruction if there is no modification to the bins. Conversely, strong bin modifications lead to distortions in restored signal because of sub-band aliasing. To reduce this effect, an overlap of the consecutive FFT windows should be used, usually between 50 and 75%. It allows increasing time resolution while keeping the chosen frequency resolution determined by the window size. It is especially important if we have chosen longer widows to get enough resolution for low frequencies which helps to better distinguish between signal and noise components.

The noise reduction function describes the way a real gain is applied to each bin. There are different methods described in literature. The most popular are: the Wiener suppression rule [3], the power subtraction, and the spectral subtraction [4, 8, 9]. All these methods apply gain reduction depending only on the power level of the corrupted signal measured at the particular bin. In our denoiser we use a parametric implementation [5] that allows adjusting any of these characteristics and in addition intermediate settings, dependent on the application. We called this parameter *Ratio*.

The most critical problem, especially if we restore high-quality music recordings, are distortions referred to in literature [10] as 'musical noise'. The reason is that low-level signals practically cannot be separated from the noise. The system cancels not only noise, but also the desired signal. It results in short tones randomly distributed over the whole frequency range. Various methods have been developed to reduce these phenomena: starting with over-estimation of the noise power spectrum, through different method of spectral magnitude averaging to intelligent time-smoothing rule of the Ephraim and Malah [10, 11] further modified in [5, 12, 17, 18]. In our work we tried to create a parametric combination of the most effective methods dependent on applications (music, speech, more artistic, more intelligible). Finally, we carefully adapted principles well known in the design of high-quality dynamics (expander, compressor) using time-varying threshold functions and proper time smoothing for every bin. The result is NoiseFree, a denoising tool implemented as a DirectX plug-in.

## 4. DENOISING PROCESSOR

### 4.1. Overview

NoiseFree effectively removes broadband noise from any kind of digitalized audio material. It has been designed to cover a wide range of applications, from the smooth high-quality denoising of musical material to the aggressive treatment of critical forensic material. Typical tasks for NoiseFree include removal of tape-hiss, surface noise of old records, broadcast noise, microphone and preamp noise, as well as enhancement of conversations, interviews and telephone cuts that lack intelligibility.


Figure 2: Screenshot of the denoising tool, NoiseFree.

The denoising process can work based on either the built-in white noise profile or a recorded noise profile. The *Learn* algorithm allows recording a noise profile from virtually any part of audio material, even if there is no noise-only part available. A 5-band *Noise Profile EQ* is provided for modifying the shape of noise profiles and smoothing them for optimal performance. The *Ambience* recovery and *Decorrelation* parameters allow fine-tuning NoiseFree to distinguish between noise and sharp signal transients. Proper *Response* parameter setup allows reduction of artifacts. The *Chase* function--mainly designed for speech denoising--automatically detects any fluctuations in background noise and adjusts the *Noise Profile* according to these changes.

NoiseFree works impeccably with sampling rates up to 384 kHz. Thus it is also perfectly suitable for hi-resolution DSD post-production. Since the CPU requirement for NoiseFree is reasonably low, all parameters can be optimized while listening to the audio material in real-time.

### 4.2. Using Denoiser

In a single-ended noise reduction system, the user decides which noise characteristic (or noise profile) has to be applied for the denoising process. NoiseFree offers two possibilities for generating the noise profile: (1) by capturing it from the input signal, (2) by creating it from flat (white) default noise. To receive the highest system performance, it is recommended to record your own noise profile from a portion of recording containing the background noise only (see later *Learn from Noise*). If the noise-only part is not available, NoiseFree provides a special mode that allows extracting the noise profile even from the complete input signal as it is, thus containing desired signal, too (see later *Learn from Signal*). After getting some experience, excellent results can also be achieved by creating the noise profile just from white noise by using a complex noise profile modifier (*Noise Profile EQ*) included in NoiseFree. Of course, recorded noise profiles can also be modified by the *Noise Profile EQ*.

When working with critical audio material, it is helpful to record (or create) a few different noise profiles and save them. Since the entire process runs in real time, we can even load the profiles during playback and listen to the result. This helps to discover which noise profile is best suited to the audio material part being processed.

The noise reduction process in NoiseFree PlugIn is controlled basically by just two parameters, *Threshold* and *Ratio,* allowing an easy search for optimal results for any given input signal. The remaining parameters are for fine-tuning. To get the best possible results in a short time we introduced a *Difference* button. It switches between the processed signal and the input/output difference, i.e., the portion of the signal being removed by the denoising algorithm. This differential signal normally should <u>not</u> contain any parts of the original signal you want to preserve.

For optical examination of the frequency spectra, a signal analyzer is provided. It displays the input signal (red) and output signal after processing (green), as well as the noise profile (white) applied to the processed audio material. We can intuitively follow the effect of the denoising process on the input signal. The noise profile (white) marks the threshold border, above which no noise reduction is applied. The *Threshold* parameter moves this noise profile up and down and can help to exactly place the profile just above the background noise level. For a given *Threshold,* the second parameter, called *Ratio,* controls the amount of the reduction applied to spectral components being below the chosen noise profile.

Different host editor setups and DirectX buffering can delay the audio path by different time intervals. In order to synchronize the *Analyzer* display with the signal being played back, an adjustable delay is provided.

A good starting value for *Threshold* is to set the noise profile just above the background noise level (approx. 10 dB). A subsequent increase of the *Ratio* parameter should significantly reduce the background noise. If noticeable artifacts in the form of ‘musical noise’ appear, decreasing the *Ratio* parameter and increasing the *Threshold* level (up to about 30 dB above the background noise) usually helps. Further reduction of artifacts can be achieved by a careful setting of the third slider *Ambience* and potentiometers in the *Expert Parameters* group.

The overall performance of NoiseFree is significantly dependent on the proper pre-selection with the one of five buttons in the *Type* group. They change the internal system resolution and rescale some internal parameters responsible for the dynamic behavior and artifact suppression. The names (*Music1, Music2, Music3, Speech1*, and *Speech2*) already suggest two application groups: music denoising and speech denoising. *Music1* is recommended for classical music without fast transients, *Music2* for average types of music, and *Music3* for percussive music including a lot of fast transients. *Speech1* is recommended for normal and forensic speech processing and *Speech2* for speech denoising and de-reverberation, especially if using *Chase* function (see later).

### 4.3. Capturing, Creating and Modifying Noise Profiles

The noise profile should ideally represent the frequency distribution of the noise to be removed from the noisy input signal. It is a kind of a spectral reference horizon used by the denoising algorithm. As previously explained, the position of the noise profile relative to the input signal can be controlled with the *Threshold* parameter. The *Analyzer* window provides an intuitive visual control of the input signal spectrum, the noise profile, and resulting output signal.

The quality of the entire denoising process is significantly dependent on the ‘quality’ of the applied noise profile, i.e., how exact the noise profile mirrors the characteristic of the noise we want to remove. In NoiseFree there are two methods for getting a noise profile: capturing it from the input signal or creating it by modifying white noise with the *NoiseProfile EQ*.

Normally, the best results can be achieved by extracting the noise profile from a part of audio material containing only the noise components to be removed. It can be the beginning of a vinyl or tape recording. It can also be an intermediate recording part that does not contain the desired signal, but only unwanted environmental noise. The recording of the noise profile from the noise-only signal should be done using *Learn from Noise* mode. The noise-only part should be at least 3 seconds long. If it is impossible to find a long enough noise-only portion we can try to set up a repetitive loop in the editor. If a noise profile includes spectral components of the signal to be recovered, they will also be removed or at least lowered in the denoising process. Therefore, much care and sensitivity is recommended when preparing noise profiles.

If we have problems finding a representative noise-only part in your input signal, NoiseFree offers the *Learn from Signal* function. This procedure extracts characteristics of the background noise from any part of the input signal, also those containing the desired signal. The length of the analyzed signal has to be at least 30 seconds. Longer parts or even the whole piece is recommended. It increases the quality of the extracted noise print. Of course, this method can capture only the background noise. Therefore in case of tape hiss, both *Learn from Noise* and *Learn from Signal* methods deliver comparable results. In case of environmental noise removal, recording noise print from a noise-only part of the input signal is the proper method.

If the noise characteristic varies strongly within the recording to be denoised, we recommend recording a few local noise prints and applying them respectively. To simplify the denoising of recordings with a strongly changing noise level, an automatic noise-chasing mode has been implemented. It continuously readjusts the noise print during the playback dependent on the audio material. It works well on broadband background noise and for different kinds of speech recordings. It is useful in real-time applications, e.g., denoising speech coming over radio receivers. Turning the *Chase* potentiometer to the right activates this mode. Higher numbers provide faster re-adjustments of the noise profile, but can cause artifacts. When working with true stereo signals, it is recommended to open *Advanced Options* dialog and try *Source detection* mode (see later). It allows better signal-to-noise separation. To get stable initial conditions for beginning of a denoised track in the *Chase* mode, an initial noise profile should first be recorded with one of the two previously described *Learn* methods.

Modifications of the noise profile made by using the *Noise Profile EQ* can help to get better denoiser performance. In the case of hiss, just the high-shelf modifier can help in finding a proper balance between removed noise and desired ambience. In critical cases more modifiers should be applied to increase/decrease denoising of certain spectral components and reduce the amount of artifacts. The *Noise Profile EQ* also includes a *Smoothing* function that can flatten the noise profile. Smoother noise profiles normally introduce fewer artifacts, but may decrease the transparency of the signal.

The third method of generating noise profiles is to create them by modifying the white (flat) noise spectrum with the *Noise Profile EQ*. After getting some experience with real (recorded) noise prints it is quite easy to create a correct artificial print.

The current noise profile can be stored for later use in the *Noise Print EQ* window.

### 4.4. Expert Parameters

When increasing the intensity of the denoising process, artifacts in the form of ‘singing birds’ or ‘musical noise’ (whistle-like tones) can occur. Provided a correctly captured (or created) noise profile and an optimal adjustment of *Type*, *Threshold*, and *Ratio,* artifacts can be effectively minimized by proper setup of *Ambience* slider and potentiometers in the *Expert Parameters* group.

As mentioned before, one of the most serious problems for the denoising algorithms is difficulty in distinguishing between unwanted noise and low-level signals and authentic ambience. At a high degree of denoising it results in dull sound and lack of liveliness in the output signal. To minimize this effect we introduced low-level harmonics and ambience recovery. Only one slider, *Ambience*, controls the whole process. It prevents higher harmonics to be removed when the denoising process is set up very aggressively. In some cases, too high *Ambience* settings can result in increasing undesired noise level or even in distortions in case of speech signals. For experts there are some possibilities to change a few internal parameters of the ambience recovery algorithm in a special window. We can set up the *Order* parameter saying how many harmonics related to any fundamental in the range up to approx. 1.5 kHz have to be recovered, as well as define the frequency range in which the harmonics should be recovered.

Further reduction of artifacts can be achieved by adjusting the *Decorrel*ation potentiometer. It recovers small transients normally smoothed flat in average denoising systems and keeps details in high quality music recordings alive. This feature is implemented by use of a special subtractive dithering method.

The dynamic behavior of the denoising algorithm can be adjusted by changing the *Response* parameter. It is coupled with several sub-parameters controlling time response behavior in every frequency bin. In general, higher values work well in most cases, effectively reducing the amount of artifacts. Smaller values can be advantageous if the input signal contains extremely sharp transients. However, the probability of getting more artifacts increases.

Fine adjustment of the parameters: *Ambience, Decorrel, and Response is* an iterative process requiring some experience and a good feeling for the trade-off between original signal, remaining noise, and artifacts. Often, when working with critical audio material the iterative process must again include adjustment of basic parameters and modification of the noise profile.

For forensic applications, artifacts are usually not that critical. The main goal there is to get as much valuable information as possible. For speech recordings the most important parameter is intelligibility. Therefore applied noise reduction can be much more intensive compared to the music's restoration. *Response* can normally be kept quite large, but lower values give more abrupt noise reduction. This can be advantageous with extremely noisy speech. With *Noise Print EQ* we can apply spectral modification to isolate the speech and increase *Ambience* and *Decorrel*ation for improvement of the intelligibility. An additional combination of two parametric EQ's in the signal path, first one before NoiseFree and the second one after, might help to extract important information. The first PEQ can emphasize particular spectral details before the denoising process and the second PEQ with mirror setup flattens the processing chain characteristic in order to keep the original spectral balance of the input signal.

The last parameter in the *Advanced Parameters* section is *Chase*. It controls the speed of noise print adoption to the background noise changes in the running signal. It is very useful when cleaning speech location recordings and automatic tape hiss removal. The method of signal chasing can be selected in a special menu. *Normal* works in every situation. *Source detection* mode can further improve the noise profile adaptation, but only for true stereo recordings. It reduces the influence of the desired signal sources on the noise profile by subtracting them from input signal before calculating the noise profile. *Threshold* should be set up experimentally according to the characteristics of sound sources.

NoiseFree has been developed to remove broadband noise. Therefore it should not be applied directly to audio signals containing strong impulsive disturbances like clicks or crackles. First impulsive noise must be removed by using a declicker and decrackler, e.g., our ScratchFree tool combining both functions.

NoiseFree provides an extremely useful feature that permits hearing the difference between input and restored output signal, thus the signal being removed during the denoising process. This is possible because the denoising we implemented is a linear-phase process. The difference signal tremendously helps in the proper parameter setup. As soon as we start to hear in the difference signal any parts of the material we want to recover it indicates that our setting may already be too strong. Using the *Bypass* button that allows switching directly to the input signal and with the two temporary presets [*A*] and [*B*] that allows comparing two different settings, we achieve further possibilities to optimize proper parameter setup.

### 4.5. Noise Profile EQ

If we want to modify a recorded noise profile or create an artificial one from the white noise, we open the *Noise Profile EQ* by clicking on the *EQ* button in NoiseFree main window.

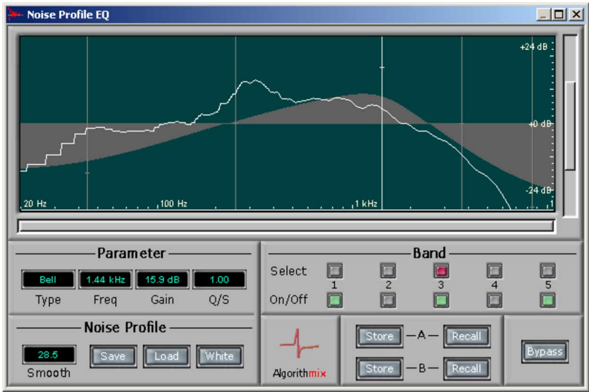


Figure 3: Screenshot of the Noise Profile EQ.

The 5-band modifier works on the noise profile like a PEQ on audio signal, e.g., if we boost a certain frequency band, the noise profile goes up in that area, resulting in more denoising. Cutting in a certain frequency range results in less denoising in that range.

A modifier band can be selected by clicking on its graphical cursor in the display or by clicking on the corresponding *Select* button. The modifier type can be chosen from three different possibilities (low-shelving, high-shelving, and bell) after opening the list by clicking on the *Type* field. The parameters of the selected band can be altered in the frequency-response characteristic, by clicking on the numeric fields or by double-clicking the numeric fields to enter a value numerically.

A noise profile previously recorded with *Learn* can be smoothed using the *Smoothing* parameter. This can be useful to remove spikes or disturbances from recorded profiles. A smooth profile will usually result in fewer artifacts, but since the profile is blurred by increased

*Smoothing*, sharp features of the signal like harmonics from hum and low frequency noise may be treated less efficiently.

The recorded profiles can be saved and loaded to/from disk using the *save* and *load* function. The *White* button erases the recently recorded profile and loads the default flat white-noise profile. For fast comparison of two *Noise Profile EQ* settings two temporary presets [A] and [B] are provided.

## 6. SOUND REPAIR PROCESSOR

### 6.1 Introduction

Denoising techniques described above work well for stationary or quasi-stationary broadband noise. Short clicks coming from old records or switching noise can be successfully removed and replaced by commercially available declicker plug-ins. However, neither a denoiser nor a declicker can successfully cope with longer complex audio disturbances overlapping material we want to recover. This fact has motivated us to develop a special tool, reNOVAtor, with an unusual in audio field interface.

Initially it was intended to rescue high-resolution live recordings from extraneous sounds and unwanted audio disturbances. Often, from the performance point of view, brilliant live recordings can not be accepted for CD release, because of disturbing noises like coughs, squeaky chairs, horn of a passing truck, a bell from the neighboring tower, or ring tones from a cellular phone. Such disturbances are especially critical if they occur in quiet passages when recording classical music. In these cases any kind of traditional equalization or editing methods is very time consuming and causes audible discontinuities in desired signal and ambience.

In most forensic recordings, the desired audio signal is even more hopelessly covered by unwanted noises. Sometimes heavy hum or buzz with many strong harmonics makes recovery of the desired signal especially difficult. Quite often we are confronted with situations where the signal-to-noise ratio is close to or even below zero.

reNOVAtor helps when all other editing tricks and processing methods fail. It allows localization, identification and very precise removal of unwanted audio events mostly without audibly affecting the audio material we want to keep. The removed sound is replaced by a signal re-synthesized from the surrounding material. Unlike time editing methods, reNOVAtor does not make deep gaps in the sound track when eradicating a disturbing sound event. Rather, it's an exactly tailored hole in the spectral representation of the processed signal that can be removed and replaced. The processing may even be restricted to a certain gain range within the selected area, which is very useful if only particular signal components need to be treated. The reNOVAtor window is fully resizable for increased accuracy and optimal compatibility with all screen resolutions.

reNOVAtor uses similar short-time analysis techniques including all the considerations previously described for denoisers. The spectral lines in the marked area can be reduced in the amplitude or completely removed. In the last case the missing parts are re-synthesized from the surrounding material. There are different ways this interpolation can be executed, namely in the time domain (horizontal), in the frequency domain (vertical), or two-dimensional. In a normal case, the horizontal or vertical interpolation are carried out from both sides. However, if one side contains strong disturbing signals, interpolation can be done from just one side. Depending on the size of the hole cut out in the spectral representation and the character of the audio signal, different interpolation techniques--well known from click or drop-out restoration--can be used: starting with linear interpolation, through Lagrange, allpass, spline to AR. The last one can deliver the best results [1, 19], but only if we know the signal model. Unfortunately, this is seldom the case.

### 6.2. Repairing Sound

At a first sight reNOVAtor's interface looks quite unusual, but after a little practice working with it becomes easy and intuitive. Later we will show a few screen shots to closely explain the idea behind the interface. First reNOVAtor loads the requested part of audio material marked in the editor and analyzes it. The result is displayed as a 3D spectrogram with time on the horizontal axis, frequency on the vertical axis and amplitude of the spectral components color-coded. After getting some experience, this 3D spectrogram representation allows a good feeling for localization and identification of different kinds of unwanted acoustical events or distortions. The spectral area of interest can be precisely marked with a resizable rectangular window. A *Play* button allows hearing different parts of the processed signal.

The spectrogram color range can be adjusted to the audio material in the *Range* field. The button toggles between three color mapping schemes: *physical* with red representing lowest energy over yellow, green, blue to white representing the highest energy; *standardized* with blue representing the lowest energy, over green, red to white representing the highest energy, and *monochrome* from black representing the lowest energy, through different gray tones to white representing the highest energy.

The marked area can be zoomed according to the menu shown in the screen shot below which opens after a right-click. In addition, it can be moved or resized as indicated by the mouse cursor. Its length in seconds and milliseconds is displayed inside. The shading grade of the area can be controlled by the *Bright* parameter from darkening (as shown below) to brightening.

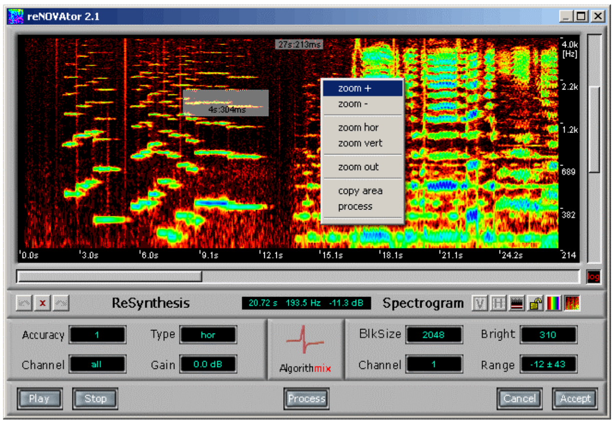


Figure 4: Screen shot of the sound repair tool, reNOVAtor.

Another way to select the visible spectrogram portion is by resizing and moving the scrollbars beside the display. By left-clicking and moving the edges of the horizontal or vertical scrollbar, the visible spectrogram region can be zoomed in or out and moved. When the SHIFT-key is also held down, zooming is performed symmetrically. The length of the displayed signal is indicated in the upper part of the spectrogram.

The displayed audio channel can be toggled by left-clicking on the *Channel* field. Block size applied to the spectrogram and interpolation is accessed by left-clicking on the *Blocksize* field, holding the button and selecting the value from the pop-up list (selectable values: 64 to 32768). In general, interpolation of short disturbances (like clicks) requires smaller block sizes, while a frequency selective interpolation (like removal of discrete tones or harmonics) requires larger block sizes.

If the mouse cursor is within the spectrogram display, the precise properties (time, frequency and amplitude) of the current cursor position are displayed in the field below the spectrogram.

The whole loaded audio material or any portion of it can be replayed at any time by positioning the white play cursor at the desired position and clicking the *Play* button.

To treat an unwanted disturbance, we first select an area around it and set up the desired parameters: *Accuracy*, *Channel*, *Type* and *Gain* (look at the first marked click below).

To remove clicks as shown in Figure 5, we usually use *horizontal* interpolation to replace the selected area containing the respective click with a new signal re-synthesized from the material surrounding the click along the time line. The interpolation types *left* or *right*

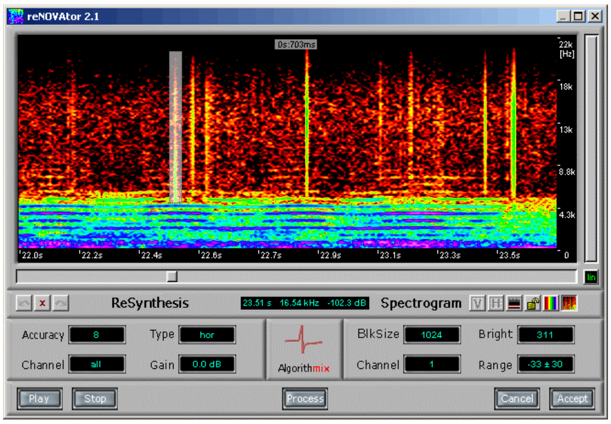


Figure 5: reNOVAtor with a marked click.

can be used, if one side of the click is not suitable for proper interpolation (e.g., it includes strong percussive beats). However, the results in case of one-side interpolation are less accurate.

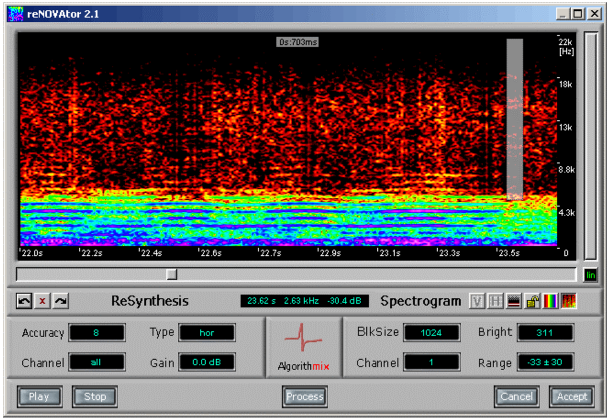


Figure 6: The same audio piece like in Figure 5, but after click removal.

To perform the interpolation in the marked area, we need to hit the *Process* button or select *process* from the drop-down menu. After processing, we can immediately listen to the result directly by resetting the *play cursor* and hitting *Play* button. The screen shot above shows the results after click removal.

If we are not satisfied with the result, we may undo the interpolation with the  button or by pressing ctrl-z. The maximal number of undo steps is limited only by the available computer memory. If we reach the memory limit we can clear the undo buffer by hitting the  button. Once an undo step is performed, it can still be re-done by hitting the  button or ctrl-y on the keyboard.

To remove discrete tones or harmonics, the *Blocksize* parameter related directly to short-time FFT size should be higher to increase the frequency resolution, especially for low frequencies. As distortions and sibilants create non-harmonic, but timely properly related tones, they can also be easy identified and removed or reduced, similar to an extremely selective linear-phase equalizer, and finally filled in with re-synthesized material. Therefore reNOVAtor can also be used as an ideal de-clipper and de-esser.

In most host audio editors, as well as in the editor included in the stand-alone version of the reNOVAtor, an additional, very useful feature is provided, namely the opportunity to preview the difference between the original and the processed signal. Using this function we can exactly hear the audio event that has been removed. If we hear parts of our desired audio material in the difference signal, we can readjust the parameter setup and marking window to get a more appropriate result.

To accomplish especially peculiar forensic tasks we have developed various advanced functions described below.

### 6.3. Advanced Functions

#### *6.3.1 Gain Selective Interpolation*

The *Gain Selective Interpolation* limits the signal repair operations to a certain gain range only. That way we can treat selected spectral components with increased selectivity, keeping as much of the original signal as possible.

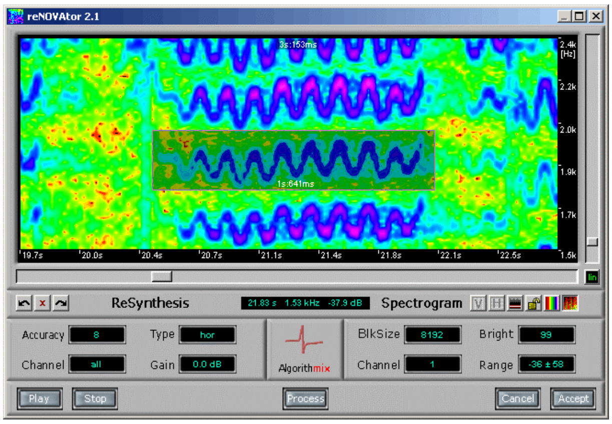


Figure 7: Harmonic to be removed marked in the reNOVAtor.

To perform an interpolation in the *gain selective* mode, we draw a marquis around the portion to be treated and hit the *lock* button. The *Range* parameter now shows the interpolation gain range rather than the overall spectrogram mapping gain.

The *Gain Selective Interpolation* applies to any selectable *Type*. The example in the screen shot above illustrates the interpolation type *gain*. As in a normal (non selective) *gain* interpolation, it reduces the amplitude of the signal components in the selected area

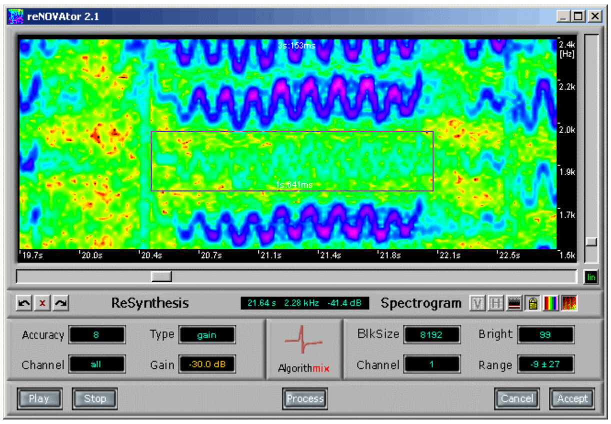


Figure 8: The same audio piece from Figure 7, but after removal of the harmonic.

(in our example by -30 dB). However, in comparison to a full-range operation, in the example above the reduction is applied only to a gain range - 40 to - 14 dB (-27 ± 13 dB). This process can be compared with a dynamic linear-phase notch filter. The dB values used in reNOVAtor for gain adjustments are related to the spectral signal component values rather than to the overall signal level.

#### *6.3.2 Copy and Paste*

While interpolation is the preferred method of removing isolated artifacts from complex audio material, there are situations where this method suffers from excess material around the artifact. In these cases (e.g., if the artifact itself is surrounded by strong transients) the *copy and paste* method, that copies an area from one spectrogram region to another, can lead to far superior results. With *copy and paste* you can simply look for another area that looks like it could fit into the space you want to interpolate and insert it.

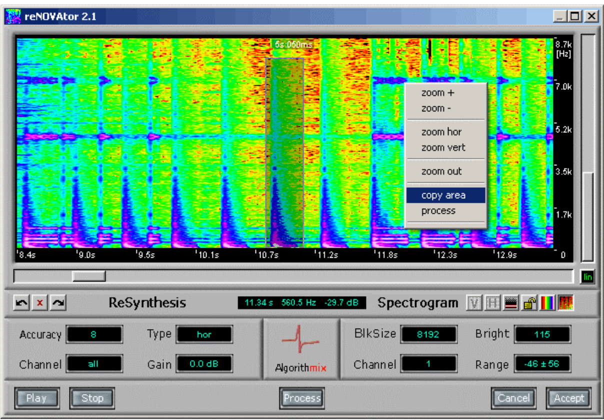


Figure 9: reNOVAtor prepared for Copy&Paste function.

*Copy and paste* can also be used gain selective (see special chapter above). This is useful for copying certain audio events or signal features (like drum beats, bells, harmonics etc.) into the time-frequency representation of the original signal, since it will copy the selected gain-range without interfering with the surrounding destination material.

Of course, this method is intended for replacing small audio parts. It helps to keep the ambience of the destination material as close to the original as possible, ensuring minimum timbral artifacts in the result. It is not suitable for replacing large pieces of audio. This is performed much better in the time-domain with an audio editor.

### *6.3.3 Automatic Selection of Harmonics*

To accelerate removal of heavy hum or buzz as well as removal or shift of complex tones including fundamentals and their harmonics, reNOVAtor provides an automatic selection of harmonics related to a selected fundamental. Of course, it would be possible to remove a complex tone just by drawing one bigger rectangular area around the whole region covering fundamental and all harmonics. For longer tones, however, strong artifacts and loose ambience integrity are expected. Using selection of harmonics we keep much more of surrounding spectrum among harmonics unprocessed. In case of continued and extremely heavy buzz this is the only method to preserve at least a part of wanted signal. It is extremely useful in forensic applications, especially if the desired signal is so strongly masked by buzz that the signal-to-noise ratio drops even below zero. Also in a case when one needs to distinguish between two or more overlapped complex tones, the <u>only</u> chance to selectively remove them is by using *Automatic Selection of Harmonics* mode.

To activate *Automatic Selection of Harmonics* mode, we need to mark the fundamental tone drawing a horizontal rectangular around it and press  button. This action pops up the *Harmonics Selection* window (see later), allowing switching on or off each individual harmonic.

This corresponds on the spectrogram to automatic drawing of a rectangle around each harmonic switched on. There are up to 19 individual harmonics and one fundamental (formally 1$^{st}$ harmonic) which can be considered in the interpolation process. The *Harmonics Selection* window monitors the level of each individual harmonic belonging to the marked fundamental. The levels of harmonics (related to their energy) can be displayed in dB or in % relative to the level of fundamental. Observing these levels you can get a good feeling about weight of each harmonic. This also helps to estimate the correct position of the rectangles marking harmonics; if only part of a harmonic bordered the displayed level is smaller.

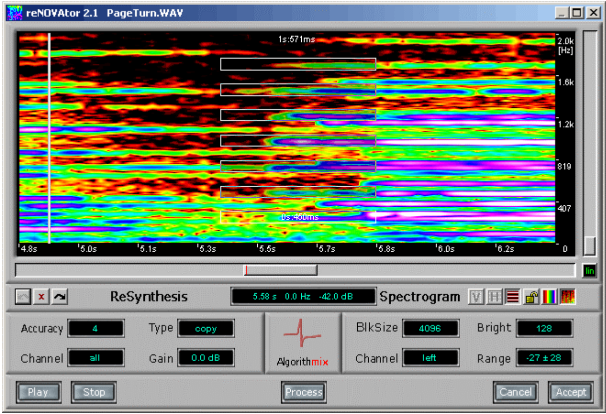


Figure 10: reNOVAtor in *Automatic Selection of Harmonics* mode.

The fundamental frequency originally selected on the spectrogram is numerically displayed in the *Freq* field of the *Harmonics Selection* window. It can be fine-tuned by left-clicking on this field and moving the mouse up or down while holding the left mouse button. This is a very important adjustment, because even small changes of the fundamental frequency extremely affect the proper position of marking rectangles for higher harmonics. Alternatively, the frequency can also be fine-tuned by holding down SHIFT while resizing the fundamental rectangle. Each individual harmonic (including fundamental) can be enabled or disabled with the respective button. Also all odd or even harmonics can be simultaneously switched on or off by using the buttons *All ODD* or *All EVEN*.

Left and right of the *Freq* field there are two additional fields. *Amp* shows the amplitude of the fundamental frequency, *Σ rms* the rms energy of harmonics (inclusive fundamental) covered by all active marked areas.

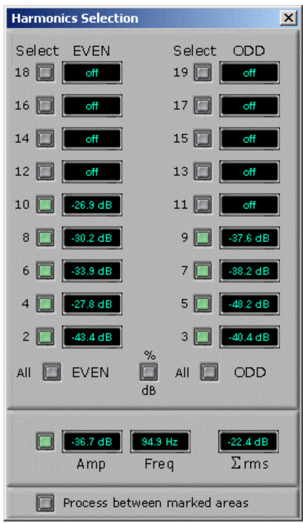


Figure 11: *Harmonics Selection* menu.

*6.3.4 Automatic Identification of Clicks*

The *Automatic Click Identification* utility automatically detects short disturbances (usually clicks) appearing as vertical lines in the marked area of the spectrogram. They can be removed all at once and the gaps re-synthesized using *horizontal* (eventually *left* or *right*) re-synthesis or just lowered using processing type *gain* (acting like a declicker). It allows automated removal of clicks from old records or spikes caused by hard switching and digital crosstalk.

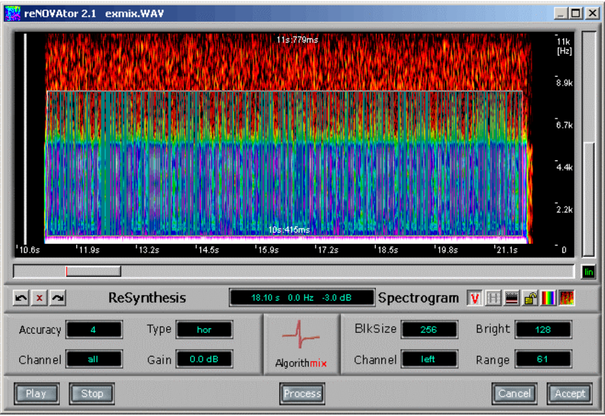


Figure 12: reNOVAtor in *Automatic Click Identification* mode.

*6.3.5 Automatic Identification of Tones and Harmonics*

*Automatic Tone and Harmonics Identification* utility automatically detects disturbances (usually tones, harmonics, hum and buzz) appearing as horizontal lines in the marked area of the spectrogram. They can be removed all at once and gaps re-synthesized using *vertical* re-synthesis type (eventually *top* or *bottom*), or lowered using processing type *gain* (acts like a linear-phase notch). It efficiently allows automated removal of very heavy hum or buzz coming from light installations like in the screen shot below.

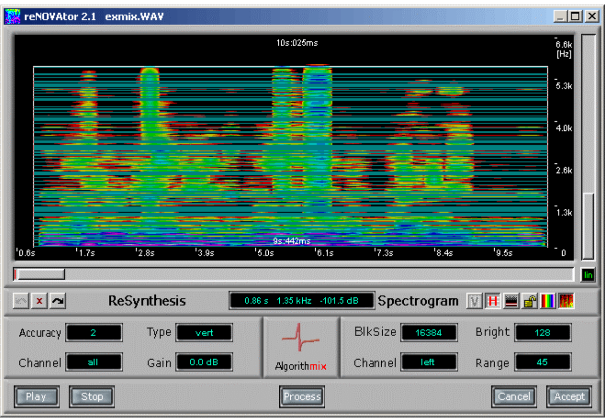


Figure 13: reNOVAtor in *Automatic Tone and Harmonics Identification* mode.

## 7. CONCLUSION

The two frequency-domain audio restoration processors, NoiseFree and reNOVAtor, described in this paper allow a new degree of quality and efficiency in audio signal restoration. reNOVAtor especially, can successfully repair signal problems which have been not solvable with any traditional audio tools. As described before, our software processors use nothing special--just well-known scientific theories. The reason for achieving superior performance in comparison to similar, previously implemented tools lies in deep understanding of psychoacoustic considerations and countless real-life tests we have performed to optimize internal parameters and couple them into well working dependencies simplifying user interface.

There are several important aspects promising a further performance increase in the sound repairing and restoration. We know well that the Fourier transform, due to its linear character regarding frequency resolution, is not optimal for audio applications. Intuitively, we would rather use non-uniform transforms or filter banks. Many experiments have already been done with wavelets. Unfortunately, the achieved audio quality in spectral-subtraction equivalent systems using strong wavelet modification between forward and backward transformation is still too weak, and/or the implementation efforts to avoid artifacts are so high that we still prefer to stay with our 'ancient' Fourier transform, despite its drawbacks.

It is said that computational power of personal computers doubles every 18 months. It invites speculation that soon we will be able to take advantage of computationally very intensive statistical methods for signal restoring and separation [1]. However, computational power isn't all that still inhibits practical implementation of these methods. There is a lack of proper mathematical models of real audio signals, an important condition for their successful use.

There have been some experiments using neural networks 'to replace' the missing knowledge about mathematical models. The results are positive, but still not as good as traditional methods, based on sound heuristic knowledge. The more details we know about physical phenomena, the more successful learn algorithms can we design.

Of course, for tools like described in this paper we would wish more intelligence in separating unwanted and wanted (ambience) noise, in a case of recordings corrupted by non-stationary noise continuously changing its level and character [14]. Also automated isolation of particular audio events or separation of audio sources based on pattern recognition and statistical methods would further help to increase the quality of sound restoration. We hope that achievements in solving similar problems in other non-audio scientific areas having more priority and thus more financial potential will also help audio community to make further progress.

For the last ten years a new trend in noise reduction based on perceptual approach has been established. It takes advantage of human auditory perception, similar to the concepts widely used in perceptual coding of audio data [13, 15, 16]. Using it in combination with previously discussed methods can further improve both the noise reduction efficiency and perceived signal quality.

Finally, we would like to point out two specific problems that should also be considered in the future research and development work on forensic audio. In today's telecommunication, analog audio already belongs to the past. The signals are usually digitized and many times coded and decoded, often with different methods before they reach the recipient. The redundancy, naturally included in analog signals, is removed (perceptual codecs) to save the bandwidth of the transmission path. Even small portable recorders, or recording functions implemented in many cellular phones are today based on perceptual coding schemas. This fact should be seriously considered in new sound restoring systems.

Another problem, especially critical in some specific applications like real-time surveillance or real-time denoising extremely disturbed telephone dialogs can be problematic, because DSP based denoisers, especially if working on extremely bad quality signals, introduce high signal latency, sometimes even longer then one second. In addition, there is often no possibility to teach the denoiser with a noise-only part, or noise character changes are so strong and rapid that we urgently need to develop new sophisticated methods to be successful in these special areas.

## 8. REFERENCES

[1] Godsill S.J., Rayner P.J.W., *Digital Audio Restoration: A Statistical Model Based Approach*, Springer-Verlag London, 1998.

[2] Godsill S.J., Rayner P.J.W., Cappé O., *Digital Audio Restoration*, in, Kahrs M., Brandenburg K., Eds., *Applications of Digital Signal Processing to Audio and Acoustics*, Kluwer Academic Publishers, 1998.

[3] Wiener N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications, Principles of Electrical Engineering Series*, MIT Press, Cambridge, MA, 1949.

[4] Boll S.F., *Suppression of Acoustic Noise in Speech Using Spectral Subtraction*, Proc. of IEEE ASSP-27, Apr.1979, pp. 113-120

[5] Etter W., Moschytz, G.S., *Noise Reduction by Noise-Adaptive Spectral magnitude Expansion*, J. Audio Eng. Soc.,Vol. 42, No. 5, pp. 341-349, May 1994.

[6] Canazza S., De Poli G., Mian G.A., Scarpa A., *Real Time Comparison of Audio reastoration Methods Based on Short Time Spectral Attenuation*, Proc. Digital Audio Effects, Limerick, Dec. 2001.

[7] Erwood A.F., Xydeas C.S., *A multiframe spectral weighting system for the enhancement of speech signals corrupted by acoustic noise*, Fifth European Signal Processing Conference: Theories and Applications, pp 1107-1110, 1990.

[8] McAulay R.J., Malpass M. L., *Speech enhancement using a soft-decision noise suppression Filter*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-28, No. 2, pp. 137-145, 1980.

[9] Vary P., *Noise suppression by spectral magnitude estimation--Mechanism and theoretical limits*, Signal Processing, Vol. 8, No. 4, pp. 387-400, 1985.

[10] Ephraim Y., Malah D., *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No. 6, pp. 1109-1121, 1984.

[11] Ephraim Y., Malah D., *Speech enhancement using a minimum mean-square error log-spectral amplitude estimator*, IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-33, No. 2, pp. 443-445, 1985.

[12] Cappé O., *Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor*, IEEE Trans. Speech, and Audio Processing, Vol. 2, No. 1, pp. 345-349, April 1994.

[13] Knagenhjelm H.P., Klein W.B., *Spectral Dynamics is more important than spectral distortions*, Proc. IEEE Int. Conf., Acoustics, Speech, and Signal Processing, May 1995.

[14] Quatieri, T. F., Baxter, R.A., *Noise Reduction Based on Spectral Change*, IEEE Workshop on Applications of Signal Proc. to Audio and Acoust., New Paltz, NY, October 1997.

[15] Wolfe P. J. and Godsill S. J., *The application of psychoacoustic criteria to the restoration of musical recordings*, Preprint 5150, 108th Convention of the Audio Engineering Society, 2000.

[16] Wolfe P. J. and Godsill S. J., *Towards a perceptually optimal spectral amplitude estimator for audio signal enhancement*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000, vol. 2, pp. 821{824.

[17] Wolfe P. J., Godsill S. J., Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement, Proceedings of the 11th IEEE Workshop on Statistical Signal Processing, pp. 496-499, 2001.

[18] Wolfe P.J., Godsill S. J., *Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement*, EURASIP Journal on Applied Signal Processing, 2003. Special Issue: Audio for Multimedia Communications.

[19] Potamitis I., Fakotakis N., *Autoregressive Time-Frequency Interpolation in the Context of Missing Data Theory for Impulsive Noise Compensation*, Eurospeech 2001, European Conf. Speech Communication and Technology, Vol. IV, pp. 1275-1278, Aalborg, Denmark, Sept. 2001.