SEMI-AUTOMATED TECHNIQUE FOR NOISY RECORDING ENHANCEMENT USING AN INDEPENDENT REFERENCE RECORDING

PAVEL IGNATOV, MIKHAIL STOLBOV, SERGEI ALEINIK

Speech Technology Center, Saint Petersburg, Russian Federation
{ignatov, stolbov, aleinik}@speechpro.com

Implementation of techniques for the enhancement of noisy recordings is an important problem for forensic applications. Widely used two-channel filtering algorithms are usually inefficient if external reference recordings are obtained from the corresponding music CD or Internet. We described new practical technique of external reference recording enhancement. This technique consists of matching the main and reference recordings using time warping of the reference signal and a two-channel spectral subtraction filtering algorithm with compensation of mismatch of the frequency response of the main and reference channels.

INTRODUCTION

Forensic recordings often have low signal-to-noise ratios caused by interfering acoustic audio signals of TV and audio devices. Enhancement of such recording is a topical issue. The conventional method of noisy recording enhancement is based on adaptive two-channel noise cancellation [1]. However, in certain situations it is impossible to record the primary and reference audio signals on a two-channel recorder simultaneously and subsequently cancel the interfering noise using two-channel adaptive filtering.

One of the methods of solving this problem is two-channel filtering using an "external" reference recording obtained from the corresponding music CD, Internet or other source [2]. But usually two-channel filtering algorithms are inefficient in this case because of mismatched conditions of recording in the primary and reference channels. The performance of these algorithms depends on a variety of factors, including effects of room acoustics, frequency characteristics and different sampling rate of audio devices, audio signal compression transform and others. These factors lead to a decrease in correlation between primary and reference signals and, consequently, to the degradation of noise suppression quality.

This paper reports a technique for enhancement of recordings corrupted by audio interfering signals which is robust to the mismatch between the primary and reference channels.

1 FUNDAMENTALS

1.1 Signal and noise assumption

Let us remark that in this paper:

- The target signal in the main (primary) channel is human speech corrupted by a noise and reverberation.
- The noise in the primary channel is the interfering sound of music and/or TV broadcasting.
- The noise (or signal) in the reference channel is the same sound of music and/or TV broadcasting, recorded in the same room under the same conditions or obtained from an external source (music CD, etc.).

Importantly, the noise in both channels is non-stationary.

1.2 Coherence and the quality of noise suppression

One of the most popular methods of non-stationary noise suppression is based on adaptive two-channel noise cancellation (ANC) [1].

Let us designate primary $x_{pri}(t)$ signal as:

$$x_{pri}(t) = s(t) + n(t) \quad (1)$$

where t is the time index, $s(t)$ is the target signal (i.e. speech), and $n(t)$ is noise in the primary channel.

The signal at the output of time-domain ANC, $y_{ANC}(t)$ may be expressed as:

$$y_{ANC}(t) = x_{pri}(t) - w(t) * r(t) \quad (2)$$

where $w(t)$ is the ANC filter impulse response, $r(t)$ is the reference signal and '*' is the convolution sign.

One of the properties inherent to the majority of adaptive two-channel noise cancellation algorithms is assuming that noise suppression quality directly depends on coherence between noises in the primary and reference channels [2]. When coherence is high, this class of algorithms is capable to suppress noise

efficiently. When coherence is low, or equal to zero, noise suppression is not realized. The precise formula linking noise coherence in channels and noise suppression ratio for time-domain ANC is given in [2].

1.3 Forensic audio recordings

Unfortunately, in real forensic audio recordings, noises in primary and reference channels are often poorly coherent. Large spacing of microphones, acoustics of the premises and mistiming of channels reduce coherence. Moreover, it often happens that the target signal in forensic audio recording represents a one-channel recording of speech mixed with the background of interfering music or TV broadcasting. In this case, there is a possibility to obtain "artificial" or "asynchronous" reference signal, e.g. from the relevant music CD, however in this case music coherence in primary and reference channels will be close to zero, because:

- The time of recording is different (no timing).
- Recording conditions are different.
- Equipment is different.
- Sample rate may be different (both strongly and with minor low drift).

In this paper we propose a process for efficient noise suppression, even in the case described above.
Let us call the proposed process a technique of two-channel "asynchronous noise reduction" (ANR).

2 PROPOSED TECHNIQUE

2.1 Basics

ANR consists of five steps:

- Obtaining the reference signal recording.
- Rough correction of the characteristics of the primary and reference signals.
- Precise alignment of the signals' beginnings.
- Precise synchronization of the signals' sampling frequency (the signals' duration).
- Denoising using two-channel noise suppression algorithm which is robust to primary and reference signals coherence.

Below we describe these steps in detail.

2.2 Reference signal

The first step of ANR is obtaining the "external" reference signal, e.g. from the relevant music CD, Internet, etc. Requirements for the reference signal are standard for two-channel noise cancellers [2]: the reference signal must repeat the "noise component" in the primary signal as much as possible. Therefore, it is preferable for the reference signal to have high quality and to be recorded in an audio format without compression, at a high sample rate.

2.3 Rough correction of signals characteristics

The second step of ANR is the correction of reference signal characteristics, making them identical to the primary signal. It is performed manually by an operator and includes:

- Correction to an equal sample rate.
- Correction to an equal average spectra.
- Rough alignment of the beginning and end of noisy fragments in the channels.

2.3.1 Correction to equal sample rate

When sample rates of the reference and main signals differ drastically (for example, 16 and 44 kHz), the sample rate must be transformed to an equal one by known algorithms. Besides, it is critical to provide frequency downsampling, rather than upsampling — i.e. (for our case) 44 kHz must be reduced to 16 kHz and not vice versa. This requirement is formulated because if we transform the lower sample rate to the higher one, the obtained signal, despite a new high sample rate, will nevertheless have no components with frequency higher than "initial sample rate / 2" (i.e. in our example above 8 kHz). This leads to differences in the reference and primary channels in the high-frequency domain. When the sample rate is reduced, there are no described above differences because the relevant high-frequency domain is cut off by the filter in the downsampling process.

2.3.2 Correction to equal average spectra

When the reference signal is taken from a CD or other high-quality source, the signal spectrum is often "richer", and high- and low-frequency components are represented more vividly compared to the primary channel. So, they will be added to the output signal during filtering. This leads to signal quality degradation. To improve the quality of noise suppression, it is preferable to convert medium (i.e. calculated throughout the entire signal length) spectrum of the reference signal accordingly so that it conforms to the noise spectrum in the primary channel to a maximum extent. This action may be performed both manually (using known equalizers, low-pass and high-pass filters, etc.) and in automatic mode using so-called algorithms of "frequency equalization", described e.g. in [3].

2.3.3 Rough alignment of noisy fragments

This step includes searching for the fragment in the reference recording, identical to the noise in the primary channel and their timing. It is required to reduce errors at the next stage of precise alignment. Rough alignment is performed manually and consists of matching the "start" and "end" tags of the interfering signal in the

reference and the main recordings. The search for the interfering fragments in the recordings may be performed both in time and frequency domains. In the first case, the operator looks for "similarity" of signal time presentation and in the second case of signal spectrograms.
It should be noted that the procedure of rough alignment may be quite difficult, since noise in the primary channel normally strongly differs from the noise in the reference channel as a result of reverberation effects and distortions produced by the recording equipment.

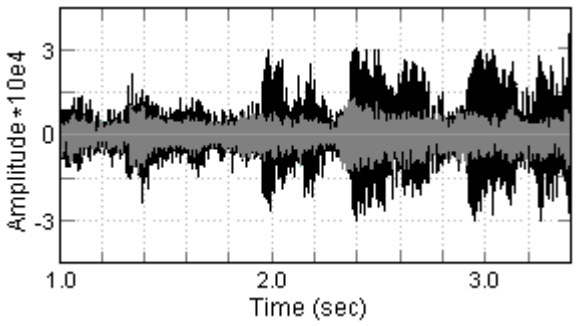

Figure 1: Primary (black) and reference (grey) signal fragments – time domain.

In this case, the operator has to listen to the recordings numerous times, view time (Fig. 1) and relevant frequency (Fig. 2) signal presentations to find out the matching time or frequency patterns, etc.

The proposed example allows us to perform alignment as a result of found initial power oscillations (Fig. 1, interval 1.0 – 1.6 s) and similar bursts in spectrograms on medium and high frequencies (Fig. 2, 2.2 and 2.9 s).

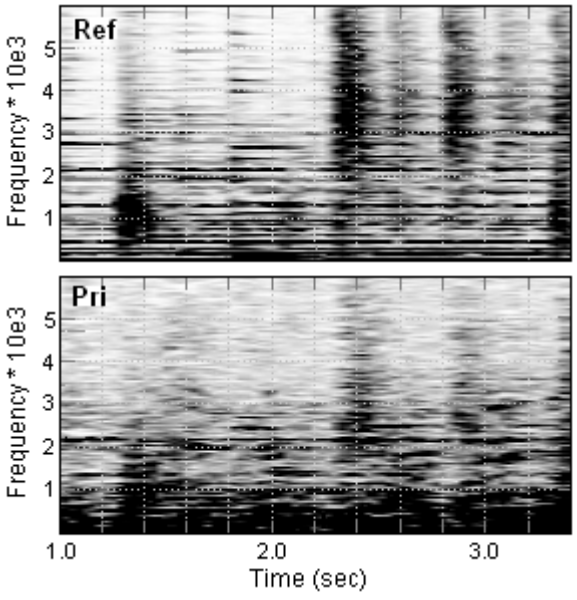

Figure 2: Primary (Pri) and reference (Ref) signal fragments – frequency domain.

2.4 Precise alignment of signals beginning

The third step of ANR is precise alignment of the noise fragments beginning in the main and reference channel.
The standard procedure of delay estimation using the maximum of cross-correlation function (CCF) [4, 5] is not effective in our case, because of low coherence between the reference and main signals.
However, our research demonstrates that apart from the signals themselves, their power envelopes are found to be sufficiently correlated. Actually, the music signal power envelope basically characterizes tempo, e.g. it mostly express power bursts of chords, percussion instruments, etc. Different channel characteristics and recording conditions retain tempo unchanged.
The algorithm for alignment of the signals beginning is presented below:

- Calculate power envelopes of the reference and main signals in frequency band.
- Calculate CCF of the power envelopes at the beginning sections of the reference and main signals.
- Find out the position of the CCF maximum and shift the reference signal to the relevant number of samples.

2.4.1 Calculation of power envelopes

It is reasonable to calculate the power envelope in a band to avoid the ambient noise, which is different in both channels. In this case, the reference and main signals are preliminary passed through a band-pass filter. The filter parameters are selected by the operator in manual mode.
Calculation of the signal power envelope by the well-known procedure consists of smoothing the signal square values using a symmetrical window, for example, Hann. The window length must be matched with the envelope variation rate (i.e. with signal tempo). During voice or music processing, (the standard tempo is within 2-16 Hz), the length of the smoothing window may be several thousands of samples.
To reduce computational cost, we have applied two-pass "to and fro" exponential smoothing, where the second reverse pass is used to compensate the phase shift of the calculated envelope. The exponential smoothing factor α corresponds to the so-called "equivalent window length" N_{eq}:

$$\alpha = 1 - 2 \,/\, (1 + N_{eq}) \qquad (3)$$

The algorithm calculating the power envelope is the following:
Let us assume that $x(i),\ i = 0, N-1$ - is the input process.
Let us assign $z(i) = x^2(i)$

In this case:

- Direct pass:

$y_{to}(0) = z(0)$
For all $i = 1, N-1$ *do:*
$y_{to}(i) = \alpha y_{to}(i-1) + (1-\alpha)z(i)$

- Reverse pass:

$y_{from}(N-1) = z(N-1)$
For all $i = N-2, 0$ *do:*
$y_{from}(i) = \alpha y_{from}(i+1) + (1-\alpha)z(i)$

- Result:

For all $i = 0, N-1$ *do:*
$y(i) = (y_{to}(i) + y_{from}(i))/2$

An example of magnitude smoothing is given in Fig. 3.

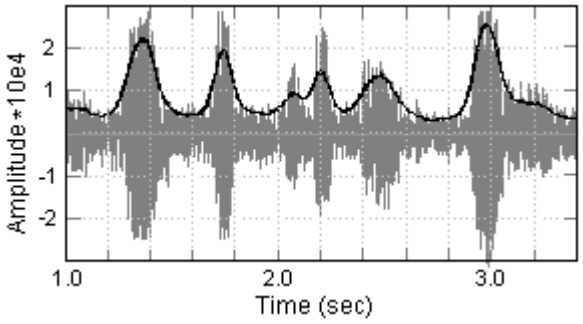

Figure 3: Initial speech signal (grey) and its "to and fro" magnitude envelope (black).

In case of music signal processing, the smoothing factor α is found from (3) for:

$$N_{eq} = 60KS_f/BPM \quad (4)$$

where *BPM* (beats per minute) is a unit typically used as a measure of tempo in music, S_f is the sampling frequency in Hz, *K* is the coefficient: $0.1 \le K \le 1$.

2.4.2 Calculation of cross-correlation function

CCF is calculated for a section of rough alignment for the beginning point of the noises in the main and reference channel. To complete the calculations for one cycle, CCF had been calculated as follows. Let us designate envelopes of the main and reference signal as $x(i)$ and $y(i)$, and the time shift as k. Further, accept $j = i - k$. Now we can write the normalized cross-correlation function (CCF) of signals power envelopes $R_{xy}(k)$:

$$R_{xy}(k) = \frac{\sum x(i)y(j) - \frac{\sum x(i)\sum y(j)}{M}}{d^{1/2}} \quad (5)$$

where:

$$d = (\sum x^2(i) - \frac{\sum x(i)\sum x(i)}{M})(\sum y^2(i) - \frac{\sum y(j)\sum y(j)}{M})$$

$M = (N-k)/L$, N is the total number of samples for CCF calculation, L is the calculation step, i.e. summing up is provided for all $i = 0,\ L, 2L, .., N-k$.
In out experiments we used a step $L = 10$, 100, ..1000 without significant loss of CCF quality.

2.5 Synchronization of noise signals in the primary and reference recordings

The fourth step of the proposed technique is synchronization, which denotes "precise sampling frequency synchronization" (PSFS).
The synchronization of the noise signals in the primary and reference recordings defines the speech enhancement capabilities of the proposed technique.
The importance of PSFM may be easily demonstrated by an actual example. In one of our experiments we discovered the following phenomenon. Because of the difference in the recording equipment, sampling frequencies of the primary and reference signals were 16000.0941 and 16000.0000 Hz, respectively. One may easily estimate that such difference resulted in the divergence of 941 samples by the period of 10 seconds. The length of the processed frame in our denoiser was selected as 512 samples. So in 5.4 seconds after the start of noise suppression, the denoiser was operating on mismatched frames. As a result, noise suppression quality degraded sharply.
PSFS algorithm includes the following steps:

- Calculation of envelopes CCF on the final primary and reference signals sections, indicated by the operator during rough tuning.
- Estimation of the delay that corresponds to CCF maximum.
- Calculation of compression/extension ratio for the reference signal:

$$\lambda = (I_e - I_b + argmax(CCF))/(I_e - I_b) \quad (6)$$

  where: I_b и I_e are the time indices of the reference signal beginning and end, respectively.
- Signal compression/extension in the reference channel.

Compression/extension in general is the procedure of interpolation of the reference signal. Linear and quadratic interpolation, tested in our experiments, demonstrated practically identical positive results.

2.6 Two-channel spectral subtraction

The last step of the proposed technique is noise reduction using 2-channel spectral subtraction (SS2).
Assume that the signal in the primary channel is described in (1), where the noise $n_{pri}(t)$ may by represented as the reference signal $r(t)$ transformed by the speakerphone and room acoustics $h(t)$:

$$n_{pri}(t) = h(t) * r(t) \quad (7)$$

(here '*' is the convolution sign).
Thus the primary signal windowed and transformed using Fourier Transform can be written as:

$$X_{pri}(f,t) = S(f,t) + H(f,t)R(f,t) \quad (8)$$

were $H(f,t)$ is an unknown transfer function, f and t are the frequency and frame indices, respectively.
Note that in our experiments Fast Fourier Transform (FFT) is performed using overlapped-windowing frames.
Spectral subtraction is the standard solution for noise reduction in stationary noise situations [2, 6].
The power spectrum of the observed signal is approximated as:

$$|X_{pri}(f,t)|^2 \approx |S(f,t)|^2 + |N(f,t)|^2 \quad (9)$$

where $S(f,t)$ and $N(f,t)$ are the spectra of the target speech signal and the additive noise, respectively.

Spectral subtraction is a frequency-based technique which obtains the clean speech power spectrum $S^2(f_k,m)$ at frame *m* for the *k-th* spectral component by subtracting the noise spectrum $N^2(f_k,m)$ from the noisy speech $X_{pri}^2(f_k,m)$, i.e.

$$S^2(f_k,m) = X_{pri}^2(f_k,m) - N^2(f_k,m) \quad (10)$$

Since the noise power spectrum is unknown, an estimated noise power spectrum $|N^{est}(f,t)|^2$ is subtracted instead:

$$|S^{est}(f,t)|^2 = |X_{pri}(f,t)|^2 - |N^{est}(f,t)|^2 \quad (11)$$

The enhanced speech spectrum is estimated using a time varying amplitude filtering of noisy signal as a product of gain function $G(f,t)$ and noisy spectrum $X_{pri}(f,t)$:

$$S^{est}(f,t) = G(f,t)\, X_{pri}(f,t) \quad (12)$$

with gain function:

$$G(f,t) = 1 - \sigma(N^{est}(f,t)^2 / X_{pri}(f,t)^2)^{1/2} \quad (13)$$

where σ is the subtraction factor.
To enhance speech under non-stationary noises it is possible to use spectral subtraction with reference signal [2, 6-9]. The primary noise spectrum $N^{est}(f,t)$ is estimated using the spectrum of reference signal:

$$N^{est}(f,t) = H^{est}(f,t)|R(f,t)| \quad (14)$$

For this purpose it is necessary to estimate the unknown factor $H^{est}(f,t)$. The coefficients $H^{est}(f,t)$ are adjusted in order to minimize the following cost function:

$$C = E[|X_{pri}(f,t)| - H^{est}(f,t)|R\ f,t\ |)^2] \quad (15)$$

where *E*[ ] is the expectation operator.
Since the transfer function is considered to be time-variable, we update it in some time intervals as:

$$H^{est}(f,t) = H^{est}(f,t-1) + \alpha(f,t)[|X_{pri}(f,t)| - H^{est}(f,t-1)|R(f,t)|] \quad (16)$$

where $\alpha(f,t)$ is a time and frequency dependent factor:

$$\alpha(f,t) = \gamma\ R(f,t)\ /(|X_{pri}(f,t)|^2 + |R(f,t)|^2) \quad (17)$$

where γ = 0.5-0.125 is the time smoothing factor.
This update is carried out when only the reference signal has sufficiently high power (i.e. more than an empirically defined value P_0):

$$P_{ref}(t) = \sum_f |R(f,t)|^2 > P_0 \quad (18)$$

with frequency range *F* = [100-3000] Hz.
Proposed algorithm related to the one described in [8].
To reduce "musical" noise the reference and signal spectra are smoothed. The smoothed short time magnitude spectra $X_{pri}^{est}(f,t)$, $R^{est}(f,t)$ are estimated using first-order IIR low-pass filters:

$$X_{pri}^{est}(f,t) = (1-\gamma)X_{pri}^{est}(f,t-1) + \gamma|X_{pri}(f,t)| \quad (19)$$

$$R^{est}(f,t) = (1-\gamma)R^{est}(f,t-1) + \gamma|R(f,t)| \quad (20)$$

The smoothed estimate of the noise spectrum in the primary signal channel is calculated as follows:

$$< N^{est}(f,t) >= H^{est}(f,t)R^{est}(f,t) \quad (21)$$

In order to obtain a noise reduction method that takes into account the varying noise level, two factors are introduced: the "over–subtraction factor" σ and the "flooring factor" G_{min}.
The estimates of the signal and noise spectra are combined to calculate the gain function according to the

magnitude subtraction algorithm:

$$G(f,t) = \max\{G_{min}, (1 - \sigma < N^{est}(f,t) > / X_{pri}^{est}(f,t))\} \quad (22)$$

G_{min} and σ are determined by subjective criteria.
To suppress a significant part of the noise and to produce comfortable ("white") residual noise, a time and frequency dependent spectral floor was used. The spectral floor is determined as follows:

$$G_{min}(f,t) = \min\{1, G_{min} N_{aver}^{est}(t) / < N^{est}(f,t) >\} \quad (23)$$

where

$$N_{aver}^{est}(t) = \frac{1}{\Delta F} \sum_f (N^{est}(f,t)) \quad (24)$$

The algorithm is realized using a Hann window with 50% overlapping. The output signal is calculated using an overlap – add procedure.

3 EXPERIMENTS

3.1 Experimental data and method

We have tested the proposed technique under three types of noise: music, speech, and pink noise.
The initial experimental data consist of three different elementary mono audio signals: a recording of a song, a recording of human speech and artificially generated pink noise. Each elementary signal (1.5 min length with sampling frequency 16 kHz) was concatenated sequentially into a special single playback file.
This file was played back in a room with the size of 6 x 5 x 3 meters and reverberation time equal to 480 msec.
Recording was provided synchronously for 2 microphones. Besides, during the first session of tests, both microphones were placed at a 1m distance from the emitting speakerphone and at a 30cm distance from each other. In the second, the third, and the forth sessions, , the microphone of reference channel remained at the same position and the microphone of the primary channel was moved away from the speakerphone to a new distance: two, three, and four meters, respectively (Fig . 4).

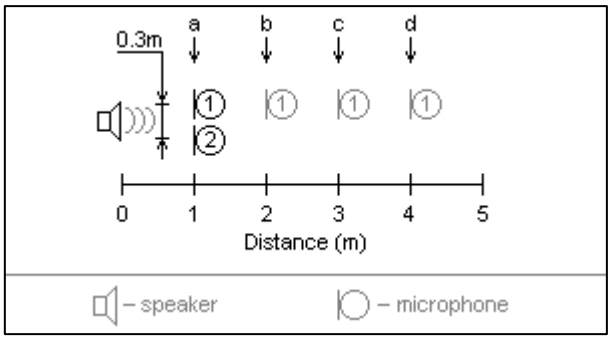

Figure 4: Arrangement of loudspeaker and microphones. 1 and 2 – primary and reference channel microphones. 'a', 'b', 'c', and 'd' – positions of the primary channel microphone in the first, second, third and fourth experimental sessions respectively.

Noise suppression in the primary channel was performed with three types of signal in the reference channel:

- Type 1: Signal recorded synchronously via the second microphone.
- Type 2: Initial test signal with synchronization of the signals' beginnings only.
- Type 3: Initial test signal with the full synchronization procedure (i.e. beginning synchronization and PSFS).

To estimate the noise reduction performance, the segmented level of noise reduction R_{seg} [6] was calculated:

$$R_{seg} = \frac{1}{K} \sum_{k=1}^{K} 10 log_{10}(R_k) \quad (25)$$

where:

$$R_k = \sum_{i=1}^{M} x^2{}_k(i) \; / \; \sum_{i=1}^{M} y^2{}_k(i) \quad (26)$$

where: K is the number of segments and M is the segment length, $x_k(i)$ and y_k i are the input and output signals at the segment k and time lag i respectively. The segment length was 40-ms.

3.2 ANC algorithm

As an adaptation algorithm we used the Normalized Least-Mean Squares (NLMS) algorithm (formula 11 in [2]) with a minor change: during signal power calculation we used exponential moving averaging. ANC filter length (N_{fil}), adaptation factor μ power smoothing factor α were varied during the experiment.
Figure 5 shows the results of noise suppression by ANC with the parameters: N_{fil}=512, μ=0.1 and $\alpha = 1 - 1/N_{fil}$ for different reference signals.

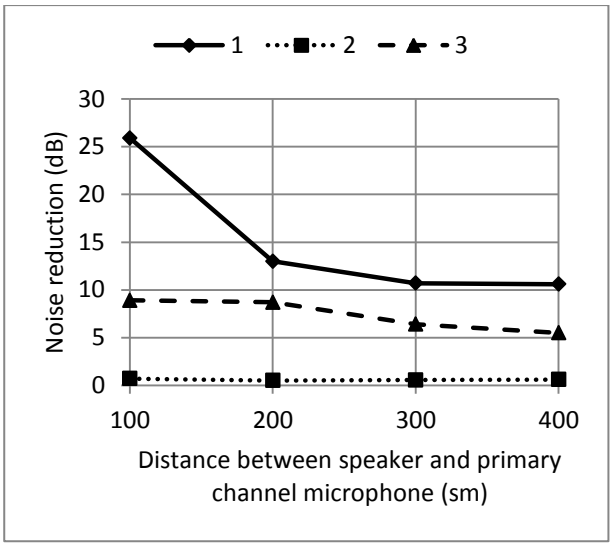

Figure 5: Segmented noise reduction level for ANC.

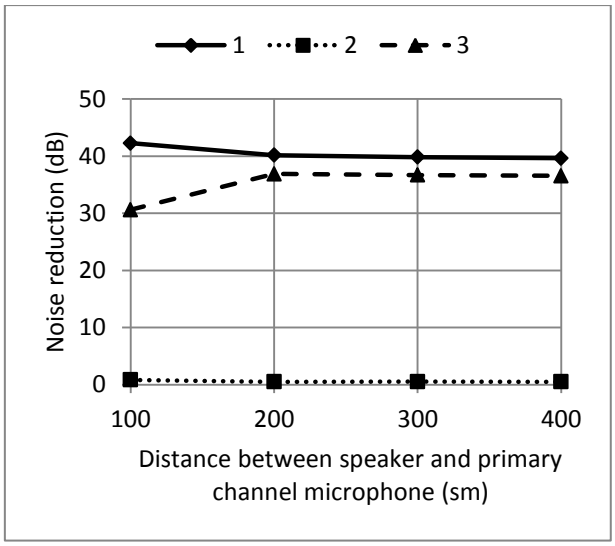

Figure 6: Segmented noise reduction level for SS2.

Curve numbers correspond to the types of reference channel signals as described in section 3.1

Curve 1 shows that ANC efficiently reduces noise level when the distance between microphones is negligible. Nevertheless the efficiency of noise suppression is reduced when the distance between microphones increases as a result of correlation drop between primary and reference signals.

However, curve 2 demonstrates that without synchronization, ANC is absolutely ineffective: R_{seg} in out tests varied within 0.3…0.6 dB.

On the other hand, curve 3 demonstrates that PSFS enhances ANC efficiency, although the observed noise reduction is not so high as in case of synchronous noise suppression.

It should also be noted that R_{seg} drop caused by the distance increase between the microphones in case of asynchronous filtering is not so clearly expressed as in case of synchronous filtering.

3.3 Two-channel spectral subtraction

The algorithm parameters are *N=512*, G_{min} *= -40 dB*, and γ*=0.25*. Curve numbers in Fig. 6 also correspond to the types of reference signal selection, as described in section 3.1.

The experiment demonstrated the following:

1. Under conditions of reverberation, the efficiency of SS2 algorithm is higher than ANC (curves 1 and 3).
2. Without synchronization, SS2 algorithm is absolutely inefficient (curve 2).
3. The synchronization procedure yields a positive effect, a minor reduction of algorithm efficiency in case of close microphone distance (curve 3, distance=100 sm) is most probably caused by the clipping of the signal from the primary channel.

4 CONCLUSIONS

This paper presented a new practical semi-automated technique of speech recordings enhancement for forensic purposes. The technique is based on the use of "external" reference recordings obtained from external sources (music CD, Internet, etc.).

A set of steps of the technique is described in detail.

We proposed an algorithm of precise synchronization for the primary and reference recordings.

We also described an algorithm of recording enhancement based on two-channel spectral subtraction with the transfer function adjustment. The algorithm enhances speech recorded under reverberant conditions with a distant microphone.

The experimental results show that the proposed technique suppresses the interfering sounds up to 30 dB. The technique has been used for the real recordings enhancement. The efficiency of the technique has been confirmed by informal listening tests.

Future work on this technique will include taking into account specific properties of sound reproducing equipment and reference recordings.

REFERENCES

[1] B. Widrow and S.D. Stearns. "*Adaptive Signal Processing*". Prentice Hall, 1985.

[2] J. Bitzer, M. Brandt, "Speech Enhancement by Adaptive Noise Cancellation: Problems, Algorithms and Limits", in: AES 39th International Conference, Hillerød/Dänemark, 2010, pp. 106-113.

[3] A. Barinov, S. Koval, P. Ignatov, M. Stolbov, "Channel compensation for forensic speaker identification using inverse processing", in: AES 39th International Conference, Hillerød /Dänemark, 2010, pp. 53-59.

[4] J. Benesty, J. Chen, Y. Huang, 'Time Delay Estimation via Linear Interpolation and Cross Correlation', IEEE Transactions on Speech and Audio Processing, Vol. 12, No-5, September 2004.

[5] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay", IEEE Trans. Acoust., Speech, Signal Processing, Vol. 24, Aug. 1976, pp. 320–327.

[6] J. Yang, "Frequency domain noise suppression approaches in mobile telephone systems", Proceedings of the 18th IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-93, Minneapolis, Minnesota, 27–30 April 1993, pp. 363–366.

[7] S. Aalburg, C. Beaugeant, S. Stan, T. Fingscheidt, R. Balan, J. Rosca, "Single- and two-channel noise reduction for robust speech recognition in car", – Siemens AG, ICM Mobile Phones, Siemens Corporate Research, Multimedia and Video technology, 2002. (http://www.ima.umn.edu/~balan/2Channel_ISCAWorkshopJune2002Germany_v1.3_LatestVersion05152002.doc).

[8] Y. Nasu, K. Shinoda, S. Furui, "Cross-channel spectral subtraction for meeting speech recognition", Proc. ICASSP 2011, pp.4812-4815.

[9] G. Nokas, E. Dermatis, "Speech recognition in noisy reverberant rooms using a frequency domain blind deconvolution method", Proc. Eurospeech, 1999, pp. 2853-2856.