NOISE ESTIMATION FOR SPEECH REINFORCEMENT IN THE PRESENCE OF STRONG ECHOES

Markus Niermann, Peter Jax, and Peter Vary

Institute of Communication Systems (ind)
RWTH Aachen University, Germany
{niermann,jax,vary}@ind.rwth-aachen.de

ABSTRACT

In announcement and communication systems, clean speech is often played back by loudspeakers in the presence of local background noise which decreases intelligibility. Speech reinforcement is a technique that enhances the intelligibility by adaptively filtering the speech, usually based on measured noise characteristics. In this contribution, solutions known from literature are applied to the case of public address systems where the microphone for noise measurement and the loudspeaker are close to each other and are far away from the listener on the ground. After establishing an acoustical model, the main problems are identified to be strong echoes between loudspeaker and microphone which make conventional noise estimation impossible and lead to an unstable system. A new approach including echo path estimation and echo-aware noise estimation is proposed and evaluated by means of simulations.

***Index Terms*—** Speech reinforcement, Speech enhancement, Noise measurement, NELE

1. INTRODUCTION

This contribution addresses the topic of speech reinforcement for public address systems using the example of speech announcements at the platform of a railway station. The intelligibility is often affected due to strong and time-varying background noise, caused by passing trains, train engines etc. While the noise cannot be combated, the intelligibility can be improved by time-adaptive speech processing, which may cover several aspects like modifying the speech level or reallocating spectral power. Many conventional algorithms require knowledge on the background noise to permit optimized processing for any noisy environment. Knowledge on the noise characteristics is therefore essential and can be obtained by means of microphones at the listener's position. For practical reasons the microphones are not installed directly on the ground of the platform but farther away from the ground and very close to the loudspeaker. This increases the echo-to-noise ratio ENR at the microphone, where the (undesired) echo originates from the loudspeaker and the noise is the (desired) background noise from the ground. For a measured ratio of $\mathrm{ENR} = 45\,\mathrm{dB}$, conventional echo cancellation (EC) methods and noise trackers fail, leading to an unstable reinforcement system. In this contribution, we present an echo-aware noise estimation algorithm which successfully estimates the noise characteristics despite severe echoes.

1.1. Relation to prior work

Speech reinforcement systems such as [1, 2, 3, 4, 5, 6] enhance the speech signal in noisy environments. Most of them exploit knowledge on the local background noise. In the context of mobile phones, speech reinforcement is called *Near-end Listening Enhancement* (NELE). NELE improves the down-link speech intelligibility of mobile phones in the presence of strong acoustical background noise. In [1] and [2], the authors present an efficient optimization algorithm which maximizes the Speech Intelligibility Index (SII) [7] under power constraints. Several parameters allow to set upper bounds for the total speech power or for the power per sub-band. To evaluate the performance, the algorithm is usually restricted to not increase the total audio power. With this constraint, the intelligibility can be improved by more than 35 percentage points [8].

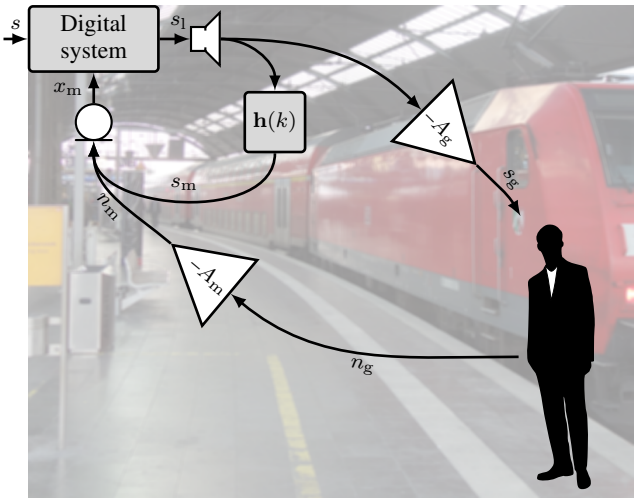

Fig. 1: Model of the acoustical part of a public address system

Listening enhancement algorithms have been adapted to several real-time applications like hand-held telephony [9] and hands-free telephony in cars [8]. The authors of [5] consider a joint optimization for different interfering playback zones, which can be applied at railway stations, for instance. Recently, intelligibility degradations due to reverberation have been integrated into the optimization [6].

Methods for estimating noise from a noisy signal, i.e., a speech signal degraded by noise, are well studied in the field of noise reduction. Possible algorithms are *Minimum Statistics* [10], *Speech Presence Probability* (SPP) [11], *Baseline Tracing* (BT) [12] and approaches based on adaptive codebooks [13, 14].

2. SYSTEM MODEL

Fig. 1 shows a model of the system. Loudspeaker and microphone are located on top of the platform, whereas the listener and the noise source are on the ground. Audio signals on the ground are indexed with 'g', signals at the microphone with 'm' and signals at the loudspeaker with 'l'. The transfer functions *microphone-ground* and

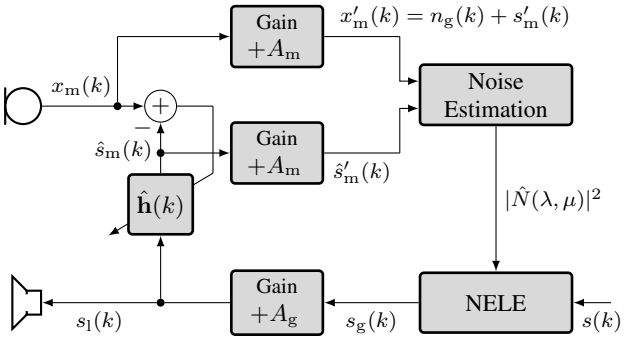

Fig. 2: Model of the digital part of a public address system with listening enhancement (NELE)

ground-loudspeaker are unknown and cannot be measured without any microphone on the ground. However, at least the level attenuation can be quantified. The corresponding attenuations A_g and A_m are known a priori and are given as positive dB-values based on offline measurements. The coefficients of the time-variant echo path between speaker and microphone are represented as the vector $\mathbf{h}(k)$ of length l_h and with time-index k. The operator $\|\cdot\|$ denotes the quadratic vector norm. The echo path may also include a path loss $A_{\mathbf{h}(k)} = -10\log_{10}\|\mathbf{h}(k)\|^2$ which for simplicity is assumed to be constant over time and therefore called $A_\mathbf{h}$. The echo path can be estimated adaptively, e.g., [15]. The estimated filter coefficients are called $\hat{\mathbf{h}}(k)$ and their quality is measured by the system distance

$$\mathrm{dist} = \frac{\|\mathbf{h}-\hat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2}. \tag{1}$$

The digital system is illustrated in Fig. 2. The NELE optimization needs the correct sound pressure levels of speech and noise on the ground because it takes psycho-acoustical effects into account. However, the system has only access to the levels at the loudspeaker and the microphone. In order to derive the levels on the ground, the path losses must be compensated for. The enhanced speech $s_\mathrm{g}(k)$ is amplified by A_g before it is played back by the loudspeaker and the recorded signal of the microphone is amplified by A_m. Succeedingly, a noise estimation is performed based on the amplified microphone signal $x'_\mathrm{m}(k)$ and also on the estimated echo $\hat{s}'_\mathrm{m}(k)$ as further information. Finally, listening enhancement is carried out using the noise estimation $|\hat{N}(\lambda,\mu)|^2$ in the frame-based frequency domain (frame index λ, frequency index μ) and the unprocessed speech $s(k)$. For the listening enhancement stage, we employ NELE [2] in a mode which allows additional audio power up to a total sound pressure level of 90 dB on the ground.

3. PROBLEM ANALYSIS

At first, the system model (Figures 1 and 2) will be compared to existing applications and afterwards, differences and emerging problems will be highlighted.

Previously, the proposed NELE algorithm has been used for hand-held telephony, e.g., without doubletalk and without coupling of microphone and loudspeaker [9]. This case is included in the current model by setting $\mathbf{h}(k) = 0$ and $A_\mathrm{g} = A_\mathrm{m} = 0\,\mathrm{dB}$. Since $x'_\mathrm{m}(k)$ contains exclusively background noise, it is sufficient to use a simple energy-based approach for noise estimation. Listening enhancement has also been used for hands-free telephony with moderate acoustic echoes between speaker and microphone [8]. Our system can be attributed to this case by setting $A_\mathrm{g} = A_\mathrm{m} = 0\,\mathrm{dB}$ and by using a room or free-field impulse response $\mathbf{h}(k)$ with $A_\mathbf{h} \geq 0\,\mathrm{dB}$. Since $x'_\mathrm{m}(k)$ contains (desired) noise and (undesired) echoes, the noise estimation block is realized in the telephone application as a conventional echo canceller with subsequent energy estimation.

In the following, a problem description will be given. In the case of public address systems, it is necessary to distinguish between the signal-to-noise-ratio on the ground (SNR) and the echo-to-noise-ratio at the microphone (ENR) due to path losses. They are related by (cf. Fig. 1)

$$\mathrm{ENR} = \mathrm{SNR} + A_\mathrm{g} - A_\mathbf{h} + A_\mathrm{m} \quad [\mathrm{dB}]. \tag{2}$$

Assuming that microphone and loudspeaker are close to and direct to each other ($A_\mathbf{h} = 0\,\mathrm{dB}$) and are located far from the ground (typically: $A_\mathrm{g} = A_\mathrm{m} = 15\,\mathrm{dB}$), ENR can be dramatically higher than SNR. Moreover, passing trains may lead to fast changing noise with high level dynamics. Also reverberation which is included in $\mathbf{h}$ makes the estimation more difficult. These effects lead to two problems, namely an unstable NELE circuit and the need for a reliable estimator of the noise spectrum despite very strong echoes.

The system may become unstable because the signal flow from NELE over $\mathbf{h}(k)$, Noise Estimation back to NELE forms a closed loop. This loop is analyzed in the following with the aid of a level model (Fig. 3). The NELE algorithm operates in critical bands (Bark scale) and modifies the clean speech s to achieve maximum intelligibility according to the SII model. The SII model predicts the maximum contribution to the intelligibility if the signal-to-noise ratio in each critical band is at least $A_\mathrm{SII} = 15\,\mathrm{dB}$. An optimum NELE algorithm would modify the clean speech s such that the signal levels in the bark bands are A_SII above the corresponding Bark noise levels of $\hat{n}$. Therefore, the NELE algorithm is represented in the level model as an amplifier which sets the level of s_g such that it is A_SII higher than the level of the estimated noise $\hat{n}$. The absolute level may be clipped with respect to psychoacoustical limits. The level model illustrates a backward influence on the noise estimation with a loop gain of $G_\mathrm{L} = A_\mathrm{SII} + A_\mathrm{g} - A_\mathbf{h} + A_\mathrm{m} \approx 45\,\mathrm{dB}$. Since $G_\mathrm{L} > 0\,\mathrm{dB}$, the speech level in s_g would rise infinitely and cause an unstable loop.

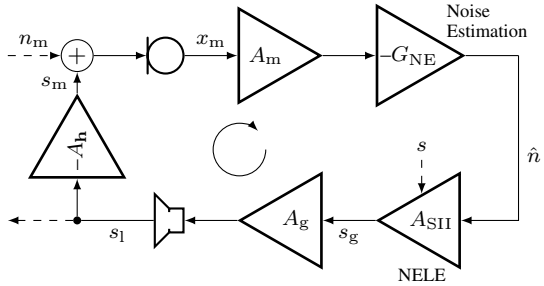

Fig. 3: Level model of the system. The gains A_m and A_g compensate for the acoustical path losses

One attempt to achieve stability could be to modify the Noise Estimation such that an echo canceller preceeds the energy measurement. It will attenuate the echo level by G_NE. For stability, the loop gain must be compensated for by the noise estimation:

$$G_\mathrm{NE} \geq A_\mathrm{SII} + A_\mathrm{g} - A_\mathbf{h} + A_\mathrm{m}. \tag{3}$$

In theory, this is possible since high ENRs improve the performance of an echo canceller. However, practical echo cancellers are not able to track the increasing echo fast enough. Therefore, we combine an echo canceller with a noise tracker to accomplish stability.

Candidate trackers are Minimum Statistics [10], SPP [11] and Baseline Tracing [12]. All of them work in the frequency domain and produce an estimate $|\hat{N}(\lambda,\mu)|^2$ of the noise periodogram which should not be influenced by speech in theory. As proposed in [11] the

estimation quality is rated by the logarithmic error distortion measure

$$\mathrm{LogErr} = \frac{1}{LM_F} \sum_{\lambda=1}^{L} \sum_{\mu=1}^{M_F} \left| 10 \log_{10} \left(\frac{|N(\lambda,\mu)|^2}{|\hat{N}(\lambda,\mu)|^2} \right) \right| \quad (4)$$

with the FFT-length M_F and L being the number of frames. Moreover, LogErrOver (LEO) rates the errors due to overestimations:

$$\mathrm{LEO} = \frac{1}{LM_F} \sum_{\lambda=1}^{L} \sum_{\mu=1}^{M_F} \left| \min \left(0, 10 \log_{10} \left(\frac{|N(\lambda,\mu)|^2}{|\hat{N}(\lambda,\mu)|^2} \right) \right) \right| . \quad (5)$$

All these noise trackers are designed for moderate signal-to-noise ratios in the range $[-5\,\mathrm{dB}, 15\,\mathrm{dB}]$, but they fail in extreme situations with $\mathrm{ENR} \approx 45\,\mathrm{dB}$ as considered here. In this scenario, speech leads to over-estimations of $|\hat{N}(\lambda,\mu)|^2$ since it is wrongly estimated as noise. Fig. 7c visualizes the over-estimation problem and Fig. 5 illustrated that this effect is increasing with ENR. Our simulations have confirmed that stability cannot be achieved.

The noise estimation gain G_NE for the stability criterion in Eq. 3 cannot be derived in a closed form. Therefore, we investigate the relationship between LEO and the power of the amplified speech s'_m (cf. Fig. 2) by means of simulations. A sufficient condition for stability is given if LEO does not increase with rising levels of s'_m. In this case, the closed loop in the level model (Fig. 3) is opened at the noise estimation.

4. PROPOSED NOISE ESTIMATION

In this section, we present a new noise tracker based on the baseline tracing approach [12] that can handle high ENRs and fulfills the stability requirement. It exploits as shown in Fig. 2 the amplified noisy microphone signal $x'_\mathrm{m}(k)$ and additional knowledge on the estimated echo $\hat{s}'_\mathrm{m}(k)$ to estimate the noise periodogram. Its general idea is to decelerate or even freeze the adaptation in time-frequency-bins where the echo level is high compared to the noise level, whereas pursuing fast adaptation in time-frequency-bins under better conditions.

The noise estimation's time and frequency resolution does not need to be as high as for applications like noise reduction since NELE averages $|\hat{N}(\lambda,\mu)|^2$ in terms of time and frequency. Moreover, time delays in the range of up to one second are not critial with respect to the perceived audio distortion. In case of too fast varying background noise the adaptation is decelerated and the intelligibility decreases temporarily, but this effect is not perceived as unpleasant.

The block diagram in Fig. 4 shows the proposed noise estimator which is used in Fig. 2. At first, an analysis transforms the input signals $x'_\mathrm{m}(k)$ and $\hat{s}'_\mathrm{m}(k)$ to the frequency domain by means of segmentation, windowing and Fast Fourier Transformation (FFT). For simplification, the time- and frequency indices λ and μ are omitted in the following. The dashed part is a conventional baseline tracer as proposed in [12]. In ordinary applications, the input of the baseline tracer is the noisy microphone signal. Here, an echo cancellation is performed beforehand such that the input $X'_\mathrm{m} - \hat{S}'_\mathrm{m}$ is composed of noise and residual echo.

4.1. Stepsize Control Rule

The stepsize control unit is the central element which enables the state-of-the-art noise tracker to measure noise in the presence of strong echoes. The baseline tracer tracks the noise by multiplying the previous estimation with a linear gain β or adding a logarithmic gain $\Delta = \ln \beta$, respectively: $\ln |\hat{N}(\lambda,\mu)|^2 = \ln |\hat{N}(\lambda-1,\mu)|^2 \pm \Delta(\lambda-1,\mu)$. Based on $\hat{S}'_\mathrm{m}$ and $|\hat{N}|^2$, the stepsize is chosen to be low if the echo is significantly louder than the noise estimation in order to freeze the adaptation. In the opposite case, when the echo level is low compared to the noise, a high stepsize permits to track the noise quickly. The baseline tracer provides the noise estimate $|\hat{N}|^2$ which is passed to NELE. This estimate leads to stability due to slow adaptations in speech periods, but noise level rises during speech cannot be tracked.

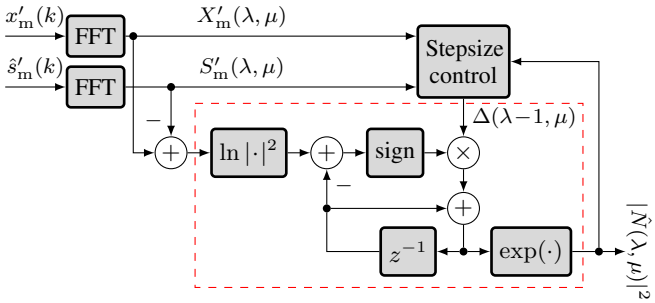

Fig. 4: Proposed noise estimator. FFT denotes the Fast Fourier Transformation. The dashed rect represents the baseline tracing algorithm [12]

To avoid freezing when the noise level increases during a speech period, a second internal noise estimation is performed:

$$|\hat{N}_\mathrm{int}|^2 = \max\left(\overline{X^2} - \overline{S^2}, 0 \right) . \quad (6)$$

$\overline{X^2}$ and $\overline{S^2}$ are obtained using a causal moving average filter (MA) of length t_MA (or an equivalent autoregressive filter):

$$\overline{X^2} = \mathrm{MA}\left(|X'_\mathrm{m}|^2 \right) , \qquad \overline{S^2} = \mathrm{MA}\left(|\hat{S}'_\mathrm{m}|^2 \right) . \quad (7)$$

The internal estimate $|\hat{N}_\mathrm{int}|^2$ alone would not meet the stability requirements, but it can detect unexpected high input levels at X'_m which result from rising noise levels. For the stepsize control, both noise estimations are combined according to

$$|\hat{N}_\mathrm{c}|^2 = \max\left(|\hat{N}|^2, |\hat{N}_\mathrm{int}|^2 \right) . \quad (8)$$

The inverse of the a-posteriori SNR γ

$$\frac{1}{\gamma} = \frac{|\hat{N}_\mathrm{c}|^2}{|\hat{N}_\mathrm{c}|^2 + |\hat{S}'_\mathrm{m}|^2} \in]0,1] \quad (9)$$

is a soft-value indicator between zero (exclusively echo, no adaptation) and one (exclusively noise, fast adaptation). Based on this, the linear tracing factor β from [12] is determined to

$$\beta(\lambda,\mu) = 1 + \gamma^{-1}(\lambda,\mu) \cdot c \cdot \frac{L_\mathrm{A}}{f_\mathrm{s}} \quad (10)$$

with the frame advance L_A and a sampling frequency f_s. The adjustable parameter c is usually chosen to allow a maximum change of 12 % in 10 ms, i.e., $c = 0.12/10\,\mathrm{ms}$.

5. SIMULATION

The proposed algorithm is evaluated by means of simulations using the parameters from Table 1. We use three noise recordings from a railway station[1] (two passing wagon trains, one departing passenger train). The recorded trains cause noise levels of up to 87 dB and high level differences. A passing wagon train lets the noise level rise from 67 dB to 87 dB, for example. 500 seconds of speech are taken randomly from the TIMIT database [16]. For the path losses, we set $A_\mathrm{g} = A_\mathrm{m} = 15\,\mathrm{dB}$ and normalize $\mathbf{h}$ such that $A_\mathbf{h} = 0\,\mathrm{dB}$.

The estimator of the time-variant impulse response needs to be adapted on-line. High values of ENR on the one hand make the noise

[1] The measurements have been carried out at the central railway station of Aachen, Germany, with the kind permission of *Deutsche Bahn AG*.

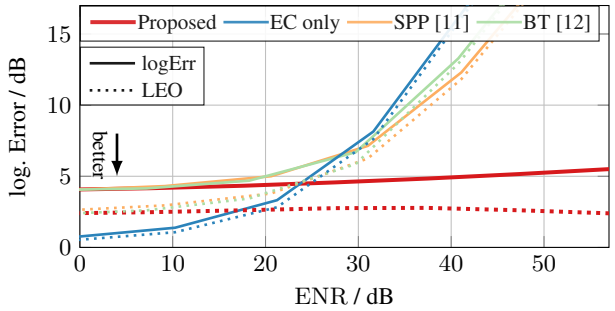

Fig. 5: Relation between log. error and echo-to-noise ratio of different noise trackers with preceding echo cancellation. dist $= -10\,\mathrm{dB}$

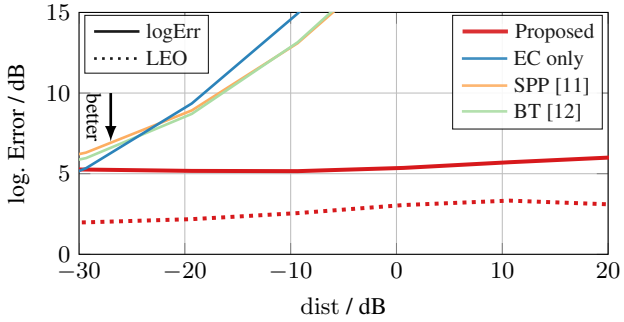

Fig. 6: Relation between log. error and system distance of different noise trackers with preceding echo cancellation. ENR $= 45\,\mathrm{dB}$, NELE disabled

estimation more difficult, but on the other hand they simplify the estimation of the echo path. It is even possible to play back perfect orthogonal sequences or sweeps during speech pauses to enhance the system identification [17]. When setting the sequence's level such that people on the ground can hardly perceive it due to masking, e.g., SNR $\approx -15\,\mathrm{dB}$, the echo-to-noise ratio ENR at the microphone, which is directly related to the resulting system distance, amounts approximately to 15 dB. This allows better estimations than typical echo cancellation applications.

In contrast, in our simulations $\mathbf{h}$ is time-invariant and $\hat{\mathbf{h}}$ is adapted once at the beginning of each noise file using a perfect sweep. Simulations with time-variant echo paths are not necessary because the system is robust to very high system distances (Fig. 6), i.e. to strong errors of the echo canceller. The impulse response $\mathbf{h}$ has been measured at the railway station mentioned above.

First, the performance of the proposed noise tracker is compared to the state-of-the-art trackers SPP [11] and Baseline Tracing [12]. An echo canceller that subtracts the estimated echo from the noisy input, i.e., $x'_\mathrm{m}(k) - \hat{s}'_\mathrm{m}(k)$, preceeds the noise trackers. The system distance between $\mathbf{h}$ and $\hat{\mathbf{h}}$ equals dist $= -10\,\mathrm{dB}$ and leads to an echo cancellation gain of 18 dB which is a realistic value for echo cancellers. The simulation results (Fig. 5) show that for ENRs higher than 23 dB, the total logarithmic error of the proposed approach is always lower or equal to the error of the reference algorithms. For ENRs below 23 dB, an echo canceller without noise tracker, i.e. $|\hat{N}(\lambda,\mu)|^2$ is derived from the error signal $x'_\mathrm{m}(k) - \hat{s}'_\mathrm{m}(k)$, performs better and is therefore the best choice in this range.

Moreover, we deduce from the simulation results that the proposed algorithm, in contrast to the reference algorithms, achieves stability. Fig. 5 shows that LEO of the proposed algorithm does not increase with ENR in the considered range. Consequently, it does not increase with rising levels of $s'_\mathrm{m}(k)$ either. According to Sec. 3, this fulfills the stability condition which is also confirmed by informal listening tests.

Parameter	**Settings**
Sampling frequency f_s	16 kHz
Frame length M	320 $\widehat{=}$ 20 ms
Frame advance L_A	160 $\widehat{=}$ 10 ms
Frame overlap	50% ($\sqrt{\text{Hann-window}}$)
FFT length M_F	512
Moving average time t_MA	200 ms
Impulse response length l_h	8000 $\widehat{=}$ 0.5 s
Path losses A_g, A_m, $A_\mathbf{h}$	15 dB, 15 dB, 0 dB

Table 1: Simulation settings

A second simulation investigates the performance for different system distances (Fig. 6). The result shows that the proposed algorithm can work with high system distances between $\mathbf{h}$ and $\hat{\mathbf{h}}$ which can occur when changes in $\mathbf{h}$ are not tracked sufficiently quick.

Thirdly, an example of one noise type combined with one speech file is visualized graphically. Fig. 7 shows a) the noise (n_g) spectrogram of a passing wagon train as well as the estimated noise periodograms when using b) the proposed algorithm and c) SPP. The estimation is performed on the basis of $x'_\mathrm{m} = n_\mathrm{g} + s'_\mathrm{m}$, i.e., noise and additive speech. While SPP still detects speech as noise, the proposed algorithm produces a realistic estimation of the noise.

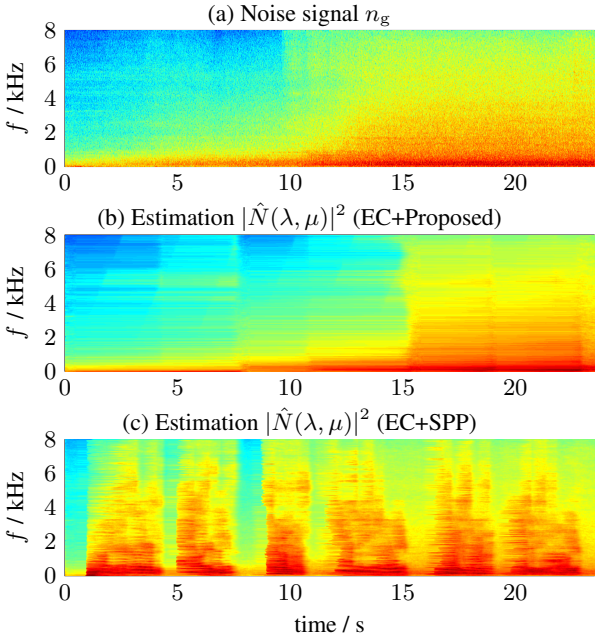

Fig. 7: Spectrogram of noise and periodograms of noise estimations. ENR $=$ 45 dB, dist $= -10\,\mathrm{dB}$

6. SUMMARY

The application of speech reinforcement with microphones and loudspeakers, which are close to each other and distant from the target area, has been analysed and has lead to the problem of noise estimation with severe echoes. To cope with the echoes and ensure stability, combinations of echo cancellation and several conventional noise trackers have been investigated. Since none of them provided satisfying results, a new echo-aware noise estimator has been developed which outperforms the regarded algorithms in terms of the logarithmic estimation error. Moreover, in contrast to the compared reference algorithms, the proposed algorithm fullfills the stability condition which has been derived in this contribution.

7. REFERENCES

[1] Bastian Sauert and Peter Vary, "Near End Listening Enhancement Optimized with Respect to Speech Intelligibility Index," in *Proceedings of European Signal Processing Conference (EUSIPCO)*, New York, NY, Aug. 2009, pp. 1844–1848, Hindawi Publ.

[2] Bastian Sauert and Peter Vary, "Recursive Closed-Form Optimization of Spectral Audio Power Allocation for Near End Listening Enhancement," in *ITG-Fachtagung Sprachkommunikation*, Berlin, Germany, Oct. 2010, VDE Verlag GmbH.

[3] T. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on power recovery and dynamic range compression," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Aug. 2012, pp. 2075–2079.

[4] Yan Tang and Martin Cooke, "Optimised spectral weightings for noise-dependent speech intelligibility enhancement," in *Interspeech*, 2012.

[5] João B. Crespo and Richard C. Hendriks, "Multizone speech reinforcement," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 22, no. 1, pp. 54–66, 2014.

[6] Richard C. Hendriks, Joao B. Crespo, Jesper Jensen, and Cees H. Taal, "Optimal Near-End Speech Intelligibility Improvement Incorporating Additive Noise and Late Reverberation Under an Approximation of the Short-Time SII," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 23, no. 5, pp. 851–862, 2015.

[7] ANSI S3.5-1997, *Methods for the Calculation of the Speech Intelligibility Index*, ANSI, 1997.

[8] Markus Niermann, Florian Heese, and Peter Vary, "Intelligibility Enhancement For Hands-Free Mobile Communication," in *Proceedings of German Annual Conference on Acoustics (DAGA)*. 2015, pp. 384–387, DEGA.

[9] Bastian Sauert, Florian Heese, and Peter Vary, "Real-Time Near-End Listening Enhancement for Mobile Phones," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. May 2014, IEEE, Show and Tell Demonstration.

[10] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 5, pp. 504–512, July 2001.

[11] Timo Gerkmann and Richard C. Hendriks, "Noise power estimation based on the probability of speech presence," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011*. 2011, pp. 145–148, IEEE.

[12] Florian Heese and Peter Vary, "Noise PSD Estimation By Logarithmic Baseline Tracing," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Piscataway, NJ, USA, Apr. 2015, IEEE.

[13] Tobias Rosenkranz, "Noise codebook adaptation for codebook-based noise reduction," in *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), Tel Aviv*, 2010.

[14] Florian Heese, Christoph Matthias Nelke, Markus Niermann, and Peter Vary, "Selflearning Codebook Speech Enhancement," in *ITG-Fachtagung Sprachkommunikation*. Sept. 2014, VDE Verlag GmbH.

[15] Gerald Enzner and Peter Vary, "Frequency-Domain Adaptive Kalman Filter for Acoustic Echo Control in Handsfree Telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140 – 1156, June 2006.

[16] John S. Garofolo and Linguistic Data Consortium, *TIMIT: acoustic-phonetic continuous speech corpus*, Linguistic Data Consortium, 1993.

[17] Christiane Antweiler, Stefan Kühl, Bastian Sauert, and Peter Vary, "System Identification with Perfect Sequence Excitation - Efficient NLMS vs. Inverse Cyclic Convolution," in *ITG-Fachtagung Sprachkommunikation*. Sept. 2014, ITG, VDE.