SIGNAL SEPARATION APPLIED TO REAL WORLD SIGNALS

*Henrik Sahlin*and Holger Broman**

Abstract

A signal separation algorithm is used in the present paper in order to improve the Signal to Noise Ratio (SNR) of a signal disturbed with noise. The algorithm uses a criterion of squared crosscorrelations between separated signals, and is thus based on second order statistics. Leaking is introduced in order to improve the performance. The signals used are real world signals measured with a modified mobile unit with two microphones. Signal to noise ratios, before and after separation, are presented. Furthermore, it is shown how to compute the SNR for signals in scenarios when both the mixing system and the noise signals are unknown.

1 INTRODUCTION

Modern telephones are very sensitive to environmental noise. The codebooks for digital communication are designed to deal with undisturbed speech, and disturbed speech can therefore become even more distorted before it reaches the recipient. For these reasons it is important to achieve a higher signal quality before the signal transmission.

The primary microphone of the mobile unit, see figure 1, registers a signal which contains both speech (referred to as the wanted signal) and unwanted noise. The subject of this paper is to present a method which attenuates the unwanted noise yielding a higher SNR value. This is achieved with a secondary microphone, in the upper end of the mobile unit, together with a signal separation algorithm.

One of the aims of this project is to collect a database of disturbed signals. The measurement equipment are a modern mobile-phone, an artificial mouth, DAT-recorders, and an anechoic room. The mobile unit is equipped with two microphones so that the signal separation method, presented below, can be used. Examples of disturbed environments are: noise played through a loudspeaker in a corner of the anechoic room, inside car driving at 90 km/h, near a drill, and in a pub.

*Chalmers University of Technology, Department of Applied Electronics, S-412 96 Gothenburg, Sweden, Phone: +46 31 772 1845, Fax: +46 31 772 1782, E-Mail: salle@ae.chalmers.se This work was financially supported by the Swedish National Board for Industrial and Technical Development.

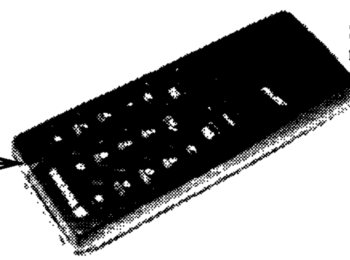

Figure 1: Modified mobile unit.

2 SIGNAL SEPARATION SCENARIO

The mixing model used in the present paper has a cross coupling structure, i.e. both source signals (x_1 and x_2) are present at both sensors (with signals y_1 and y_2), see figure 2a. The boxes denote linear, possibly time varying, dynamical filters. This problem, of separating the uncorrelated (sometimes independent) signals, is referred to as the Blind Signal Separation problem, see e.g. [2, 1]. In the context of the present paper one of these signals (x_1) is a wanted (e.g. speech-) signal and the other (x_2) an unwanted disturbance.

A separation structure is applied to the measured signals y_1 and y_2 in order to produce the signals s_1 and s_2, see figure 2b, where $D_{12}(q^{-1})$ etc. denote FIR-filters, and polynomials in the unit delay operator q^{-1}, with M_{12} coefficients of q^{-1} in the interval between M_{12}^0 and $M_{12}^0 + M_{12} - 1$:

$$D_{12}(q^{-1})y_2(n) = \sum_{i=M_{12}^0}^{M_{12}^0+M_{12}-1} d_{12}(i)y_2(n-i).$$

Note that this notation allows for noncausal filters. In the following the dependence of q^{-1}, n and the summation limits are excluded whenever possible.

The output signals from the separation structure are

$$\begin{aligned} s_1 &= y_1 - D_{12}y_2 & (2.1)\\ &= (1 - D_{12}B_{21})x_1 + (B_{12} - D_{12})x_2, \\ s_2 &= y_2 - D_{21}y_1 & (2.2)\\ &= (1 - D_{21}B_{12})x_2 + (B_{21} - D_{21})x_1. \end{aligned}$$

When $B_{12} = D_{12}$ and $B_{21} = D_{21}$ then the output signal s_1 depends solely on x_1 and s_2 on x_2, i.e. separation is achieved. Another possibility for separation is $1-D_{12}B_{21} = 0$ and $1-D_{21}B_{12} = 0$. This is referred to as the channel flop solution [3]. However, it can not occur if the all filters are dynamical FIR filters. In order to recover the signals x_1 and x_2 the output must also be post filtered with $1/(1 - D_{12}D_{21})$.

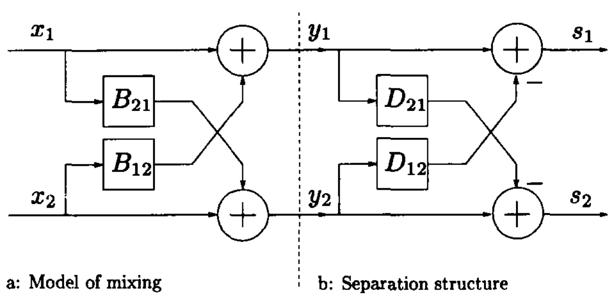

Figure 2: Two Input Two Output scenario.

2.1 Criterion based on squared cross-correlations

When separation is achieved, the two signals s_1 and s_2 are mutually uncorrelated. Thus, a criterion to be minimized can be formulated as a sum of squared cross-correlations between the separated signals [3]:

$$V = \sum_{m=-L_l}^{L_u} (\hat{R}_{s_1 s_2}(m))^2, \qquad (2.3)$$

where $\hat{R}_{s_1 s_2}(m)$ is the estimated crosscorrelation between $s_1(n)$ and $s_2(n-m)$.

The positive integers L_l and L_u are chosen equal in this paper and as $L = (Y - 1)/2$ where Y is the number of crosscorrelation values between s_1 and s_2 to be minimized. The value of Y is one of the design variables for the algorithm.

The crosscorrelations in (2.3) can be expressed in terms of the filter coefficients to be estimated and correlations of measured signals (y_1 and y_2):

$$\begin{aligned} \hat{R}_{s_1 s_2}(m) &= \hat{R}_{y_1 y_2}(m) \\ &- \sum_i d_{12}(i)\hat{R}_{y_2 y_2}(m-i) \\ &- \sum_i d_{21}(i)\hat{R}_{y_1 y_1}(m+i) \qquad (2.4) \\ &+ \sum_i \sum_k d_{12}(i) d_{21}(k) \hat{R}_{y_2 y_1}(m-i+k). \end{aligned}$$

Note that the crosscorrelation $\hat{R}_{s_1 s_2}(m)$ is nonlinear in the coefficients of D_{12} and D_{21}, which are to be estimated, but linear in the correlations of the measured signals.

The criterion (2.3) can be minimized with a Gauss-Newton search:

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \mu \tilde{\mathbf{H}}^{-1}\mathbf{G}, \qquad (2.5)$$

where $\hat{\boldsymbol{\theta}}$ is a column vector containing an estimate of the coefficients of D_{12} and D_{21}, the matrix $\tilde{\mathbf{H}}$ denotes a modified Hessian of the criterion, the vector $\mathbf{G}$ denotes the gradient and μ is a (possibly varying) step-length, see [3]. A damped Gauss-Newton search is done for each sample together with a recursive estimation of the correlations in (2.4). The algorithm is thus recursive and can be used in an on-line manner.

2.2 Criterion with leaking

In the scenario of the present paper the filters of the model, B_{21} and B_{12}, are models of acoustic paths from one of the microphones to the other. Such a path can be modeled as a pure delay (plus perhaps some echo path). Thus only a few of the parameters should be nonzero but the knowledge of which ones is limited. One way to avoid this problem is to do an order estimation of the system, i.e. to estimate how many and which of the parameters which should be used. This order estimation has to be adaptive. An alternative method is to intoduce leaking.

Another name of leaking is regularization of the criterion. It can be introduced as a modified criterion [4]:

$$W = V + \delta|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\#}|^2 \qquad (2.6)$$

where $\boldsymbol{\theta}^{\#}$ are given parameter values towards which the parameters are pulled. Without any specific knowledge of the parameters of the channels, this $\boldsymbol{\theta}^{\#}$ can be chosen to a vector with zeros, thus pulling all parameters towards zero. In a scenario with an over-parameterized model, the choice of $\boldsymbol{\theta}^{\#}$ as a vector with zeros seems plausible.

The algorithm (2.5) becomes modified with the following gradient and Hessian:

$$\mathbf{G}' = \mathbf{G} + 2\delta(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\#}) \tag{2.7}$$

$$\tilde{\mathbf{H}}' = \tilde{\mathbf{H}} + 2 * \delta\mathbf{I} \tag{2.8}$$

where **I** denotes the identity matrix.

When introducing leaking in the algorithm the variance of the estimated parameters is decreased at the cost of increased bias. The amount of leaking in the algorithm is controlled by the parameter δ. The choice of δ must therefore be done such that the decreased performance due to bias is less than the improvement due to reduced variance.

3 ENVIRONMENTS FOR MEASUREMENTS

In all of the below mentioned measurements an artificial mouth has been used. The mobile unit was mounted on the mouth at a fixed position, with the primary microphone close to the lip ring. The artificial mouth was stimulated by a sequence of voices from a TIMIT database (Texas Instruments, Inc., Massachusetts Institute of Technology). This sequence was about 2.5 minutes long and contained twelve (12) different speech signals of different dialects. Each speech signal contained two short sentences. The speakers were both male and female native English speakers.

An anechoic room (at the department of Applied Acoustics, Chalmers University of Technology) was used in order to make measurements in a controlled situation. The noise signal came from a loudspeaker in one end of the room. Measurements in the anechoic room were also made without any noise signal. These measurements were made in order to be used as the wanted signal (x_1) in the computing of the Signal to Noise Ratio. Measurements were also made inside a car driving at 90 km/h, near a drill, and in a pub.

4 SIGNAL TO NOISE RATIO FOR REAL WORLD SIGNALS

A measure of the success of the algorithm is the Signal to Noise Ratios (SNR) defined as

$$\mathrm{SNR}(y_k) = 10 \log \frac{\sum_{n=1}^{N} x_k^2(n)}{\sum_{n=1}^{N} (y_k(n) - x_k(n))^2} \tag{4.1}$$

and

$$\mathrm{SNR}(s_k) = 10 \log \frac{\sum_{n=1}^{N} \tilde{x}_k^2(n)}{\sum_{n=1}^{N} (s_k(n) - \tilde{x}_k(n))^2} \tag{4.2}$$

for k=1,2 and where $\tilde{x}_1$ and $\tilde{x}_2$ are the source signals filtered through $1 - D_{12}D_{21}$.

When separation is done with real world signals, the source signals are usually not available. In the present paper the source signal x_1 is taken as the signal recorded with the mobile unit mounted on the mouth simulator and the measurements done without any disturbances in the anechoic room. Only the values of SNR(y_1) and SNR(s_1) are computed in the present paper, since the noise signal x_2 is not availible in all the scenarios. This x_1 must be scaled and syncronized with y_1 and s_1, as described below, before subtraction in (4.1) and (4.2).

Before the sequence of the twelve speech signals is a one second sinusoidal signal. This is used in order to do a coarse syncronization of the signals. The signals are sampled at 8 kHz but in the computation of SNR a more fine syncronization is done. Before scaling, syncronization and SNR computation according to (4.1) and (4.2) the signals are interpolated to 48 kHz.

The scaling and syncronization, in the computing of SNR(y_1), are done by computing the correlation coefficients between y_1 and x_1 for a number of different lags (delays of x_1). The lag τ which corresponds to the maximal value of the correlation determines how much x_1 should be shifted. The maximal correlation coefficient ρ is used in order to scale x_1 before subtraction:

$$\mathrm{SNR}(y_1) = 10 \log \frac{\sum_{n=max(1,\tau)}^{min(N,N-\tau)} \rho x_1^2(n-\tau)}{\sum_{n=max(1,\tau)}^{min(N,N-\tau)} (y_1(n) - \rho x_1(n-\tau))^2}.$$

5 RESULTS

In the following examples, the criterion with leaking (2.6) was used. The filter which models the path of the noise source from secondary microphone to primary microphone, i.e. filter D_{12}, was given 17 taps in order to avoid constrains on the scenarios. The filter was also noncausal and had taps in the interval from q^7 to q^{-9}. The other filter which models the path from the primary microphone to the secondary microphone, i.e. D_{21}, had only 4 taps, in the interval from q^{-2} to q^{-5}. This is motivated by the location of the primary source, i.e. the mouth, which is assumed to be close to the lower end of the mobile unit, see figure 1. The distance between the microphones is approximately 13 cm. With a sampling rate of 8000 kHz this distance equals 3.1 samples (speed of sound= 332 m/s).

The values of SNR and improvements (difference between SNR(s_1) and SNR(y_1)) in table 1 are the mean of 8 speech signals. The regularization was $\delta = 10^7$, the number of equations $Y = 25$ and a for-

Sequence	mean SNR(y_1)	mean SNR(s_1)	mean improv.
180°, anechoic	-1.9	6.3	8.2
90°, anechoic	-3.1	5.7	8.8
0°, anechoic	-4.0	5.7	9.7
car	-8.1	2.2	10.4
car window	-18.1	-8.8	9.3
drill	-4.6	-0.7	3.9
pub	-11.4	-8.7	2.8

Table 1: Values of SNR before and after signal separation.

getting factor of $\mu = 10^{-5}$ was used in the estimation of the correlations for the algorithm. The first three sequences in table 1 were from recordings in an anechoic room, with the noise source positioned behind (180°), beside (90°) and in front of (0°) the mouth simulator. The two recordings in the car were made without and with one of the windows open. The noise was mainly low-frequency in these both sequences. The unwanted signal in the recordings from the drill were narrow-band and low frequency noise. The algorithm removed most of the low frequency noise but not the narrow-band signal. In the pub sequence, a lot of different sources were present. The model in figure 2 is thus not valid, i.e. the number of sources is larger than two .The algorithm removed one of the strongest of these sources from s_1.

In figure 3, the SNR and improvement are displayed as a function of time for one of the sequences. The values change with the magnitude of the speech sequence.

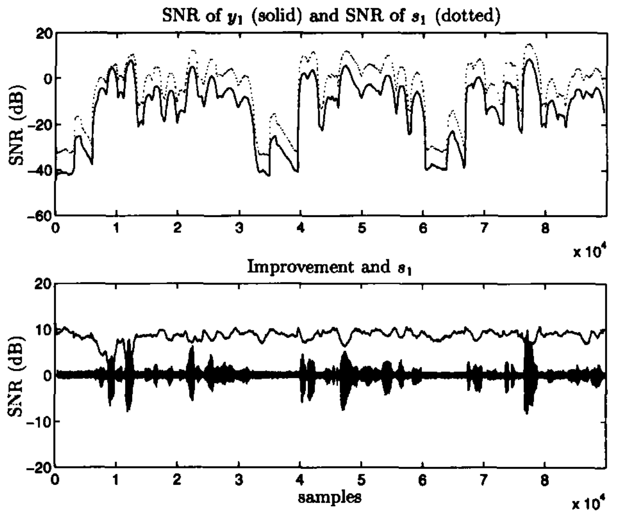

Figure 3: Signal to Noise Ratios and Improvement.

In figure 4, the final values of the coefficients of the filters D_{21} and D_{12} are shown for two sequences. For the sequence "180° anechoic" the largest coefficients are for a delay of the signal of $3-4$ samples, i.e. the coefficients for q^{-3} and q^{-4}. Compare with the sequence "0° anechoic" where the noise arrives at the primary source some 3 samples before arriving at the secondary microphone. The estimated filter D_{12} , which coefficients are plotted in figure 4, have a maximum for the coefficient of q^3.

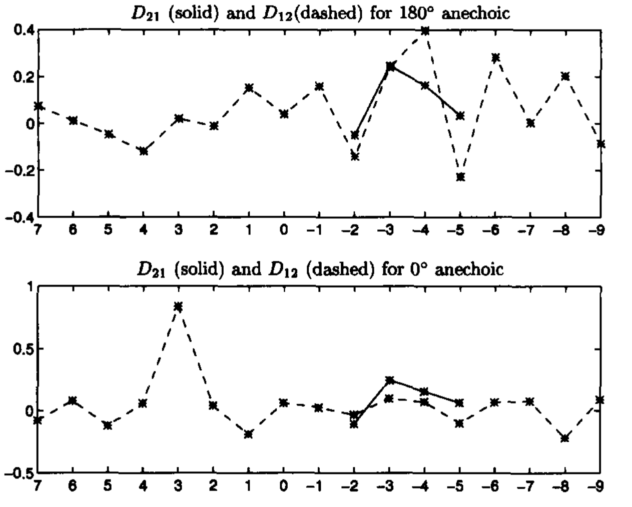

Figure 4: Final values of D_{21} and D_{12} for two signals

The signals before and after the signal separation algorithm are available at "http://www.ae.chalmers.se/~salle/eval.html".

References

[1] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, Jun. 1994.

[2] C. Jutten and J. Heuralt. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, Jun. 1991.

[3] U. Lindgren and H. Broman. Source separation: Using a criterion based on second order statistics. Technical Report CTH-TE-36, Department of Applied Electronics, Chalmers University of Technology, Göteborg, Sweden, 1996. To appear in IEEE Trans. on Signal Processing.

[4] Lennart Ljung and Jonas Sjöberg. A comment on "leakage" in adaptive algorithms. Technical Report LiTH-I-ISY-1304, Automatic Control, Linköping University, Sweden, 1992.