In: P. Pajunen and J. Karhunen (Eds.), 'Proceedings of the second international workshop on independent component analysis and blind signal separation', June 19–22, 2000, Helsinki, Finland, pp. 215–220.

AMPLITUDE MODULATION DECORRELATION FOR CONVOLUTIVE BLIND SOURCE SEPARATION

Jörn Anemüller and Birger Kollmeier

Arbeitsgruppe Medizinische Physik and Graduiertenkolleg Psychoakustik,
Carl von Ossietzky–Universität, Oldenburg, Germany
ane@uni–oldenburg.de, http://medi.uni–oldenburg.de/members/ane

ABSTRACT

The problem of blind separation of a convolutive mixture of speech signals is considered. Signal separation is performed in the frequency domain.

Based on observations from amplitude spectrograms of speech signals, the notion of amplitude modulation correlation ('AMCor') across different frequency channels is introduced. From the corresponding principle of amplitude modulation *de*correlation, a novel cost–function and an algorithm for convolutive blind source separation are derived. The algorithms' main features are discussed. Successful separation of synthetic data and of real–room recordings of speech is performed. The results of the latter are compared to the performance of previous algorithms on the same data.

Audio examples are available from the authors' web page [2].

1. INTRODUCTION

The convolutive blind source separation problem is encountered in the field of acoustics when superpositions of M source signals are recorded by N microphones in a reverberant environment. The aim is to reconstruct the source signals from knowledge of the microphone signals only. The location of sources and microphones is assumed to be unknown which gives rise to the term 'blind'.

Since sound propagation in air is linear and involves time–delays and echoes, the recorded signals $x_1(t), \ldots, x_M(t)$ are obtained as the sum of convolutions of the source signals $s_1(t), \ldots, s_N(t)$ and the room's impulse response:

$$x_i(t) = \sum_j \sum_\tau a_{ij}(\tau)\, s_j(t-\tau). \qquad (1)$$

By $a_{ij}(\tau)$ we denote the impulse response from source j to the location of the i–th microphone. We consider the case of non–degenerate source separation, i.e., when the number of microphones N is larger or equal to the number of sources M. In this case, unmixing of the microphone signals by linear filtering is feasible.

Filtering is performed in the frequency domain by first computing spectrograms of the mixed signals. Subsequently, the signals are unmixed in the frequency domain by matrix multiplication. Finally, the unmixed spectrograms are transformed back to the time–domain using the overlap–add method [10]. (See sections 2.1 and 3 for details.)

Our algorithm differs from previous algorithms by the fact that it does not try to unmix the data in each frequency channel by evaluating the mixed signals at each frequency in question independently. Rather, it integrates information across different frequencies to unmix the signals. The basis for the algorithm is the property of *amplitude modulation correlation* ('AMCor') which can be observed in, e.g., speech. It states that the the signal amplitude in different frequency bands undergoes interrelated changes (see sections 2.3 and 2.4). By applying the principle of amplitude modulation *de*correlation ('AM*De*cor') on the unmixed signals, it is possible to separate the data (see section 4).

A problem of source separation in the frequency domain are local permutations of the unmixed signals across frequency channels. They can be solved using additional constraints about the source signals [9], the room's impulse response [12, 4], or by exploiting properties of the discrete Fourier transform [14]. To this end, the amplitude modulation decorrelation algorithm comprises the inherent property that it penalizes and therefore avoids local permutations (see section 4).

Notation throughout the paper is as follows: vectors and matrices are printed in bold font; $\hat{x}_\alpha(t)$ denotes the spectrogram of quantity x; frequencies are indexed by α and β; the total number of frequency channels is denoted by Λ; the expectation operator is denoted by $E\{\cdot\}$; transposition of vector $\mathbf{x}$ is denoted by $\mathbf{x}^T$.

2. AMPLITUDE MODULATION CORRELATION

In this section we briefly review the spectrogram and the amplitude spectrogram. Motivated by observations from speech signals we introduce the notion of amplitude modulation correlation.

2.1. The Spectrogram

The spectrogram is a standard time–frequency representation used for the analysis and filtering of speech signals [10]. It incorporates spectra of short, typically 40msec long, overlapping frames of the signal. A windowing function $w(t)$ is applied prior to spectral analysis in order to enhance spectral resolution and avoid circular aliasing. The short–time

spectrum $\hat{x}_\alpha(t_0)$ corresponding to signal $x(t)$ is defined as

$$\hat{x}_\alpha(t_0) \equiv \sum_{k=0}^{K-1} x(t_0+k)\, w(k) \exp\left(-\frac{2\pi i}{K}\alpha k\right).$$

Time t_0 denotes the start of the frame, and α denotes the frequency channel. Since the complex valued spectrogram allows for transformation back to the time domain, e.g., by the overlap–add method [10], it is frequently used for digital filtering.

2.2. The Amplitude Spectrogram

By preserving only the spectrogram's amplitude and discarding the phase–information, the amplitude spectrogram $|\hat{x}_\alpha(t_0)|$ is obtained, where $|\cdot|$ denotes the magnitude of a complex number. The amplitude spectrogram is highly useful for visualization and analysis of speech signals, and it reveals the rich semi–deterministic structure present in speech. However, since phase information is lost, the amplitude spectrogram does not allow for time domain reconstruction of the underlying signal.

2.3. Structure in Speech

Fig. 1 displays the amplitude spectrogram of a speech sample. Note that many elements of this image change smoothly over both time and — more important for our purpose — frequency, and that even distant frequency channels exhibit related changes in amplitude.

The most prominent feature of speech is the amplitude modulation due to the succession of different phonemes as constituents of speech. It results in quasi–periodic, broadband maxima and minima at a modulation frequency of typically four Hertz. Another feature of speech are the spectral maxima and minima of the glottis excitation; they manifest themselves in the amplitude spectrogram's horizontal lines. The glottis is the main energy source for speech production and emits a broadband sound with spectral peaks at the harmonics of the speaker's pitch frequency. In the next stage of speech production, this broadband sound is filtered by the vocal tract which embosses the spectral shape of the phoneme in question on it. The vocal tract is a smooth physical system with, in practice, a limited number of degrees of freedom, and its transfer function is a smooth function with a relatively low number of spectral maxima and minima. Most prominent among the spectral peaks are the *formants* which are regarded as the characteristic elements of vowels [13]. E.g., the vowel [e] is characterized by simultaneous spectral peaks at frequencies of typically 500Hz and 2000Hz.

2.4. Amplitude Modulation Correlation

From the above it should be obvious that the way human speech is produced naturally leads to a similar time course of amplitude in different and even distant frequency channels. Since this can also be termed interrelated amplitude modulation in different frequency channels, we call this property *amplitude modulation correlation* [3].

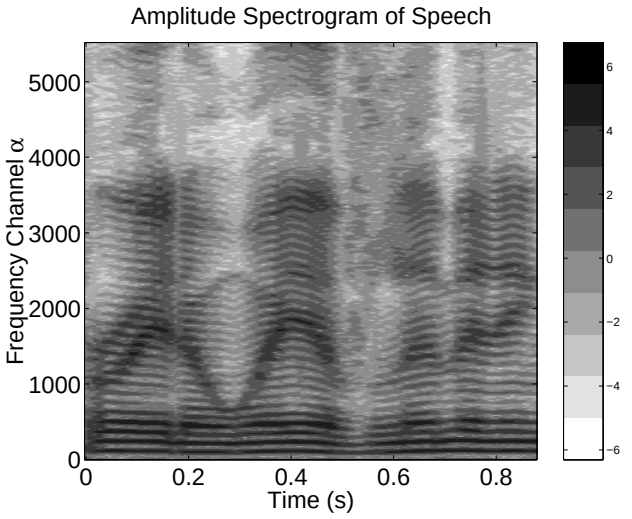

Figure 1: Amplitude spectrogram of a speech signal. From the graph it is visible that changes in amplitude in different frequency channels are not independent but interrelated, hence the term *amplitude modulation correlation*.

In order to quantify amplitude modulation correlation across two frequency channels α and β, we regard the amplitudes $|\hat{x}_\alpha(t)|$ and $|\hat{x}_\beta(t)|$, respectively, as two time series and compute their covariance. This results in the amplitude modulation correlation ('AMCor') which is defined as

$$\begin{aligned} c(\hat{x}_\alpha, \hat{x}_\beta) &\equiv E\left\{\xi^x_\alpha(t), \xi^x_\beta(t)\right\}, \\ \xi^x_\alpha(t) &\equiv |\hat{x}_\alpha(t)| - E\left\{|\hat{x}_\alpha(t)|\right\}, \\ \xi^x_\beta(t) &\equiv |\hat{x}_\beta(t)| - E\left\{|\hat{x}_\beta(t)|\right\}. \end{aligned}$$

If we compute the AMCor for each pair of frequencies (α, β) of a *single signal*, we obtain the AMCor matrix $\mathbf{C}^{xx} \equiv \left[c^{xx}_{\alpha\beta}\right]$ whose (α,β)–element is given by $c^{xx}_{\alpha\beta} \equiv c(\hat{x}_\alpha, \hat{x}_\beta)$.

A typical AMCor matrix is displayed in Fig. 2. As expected, particularly high values of AMCor are reached for nearby frequencies (i.e., near the diagonal), and high values of AMCor can also be found for frequencies which are quite distant.

The crucial question for the applicability of AMCor to blind source separation is whether the AMCor computed *across two sources* vanishes. To this end, we compute the covariance between the amplitude time series $|\hat{x}_\alpha(t)|$ at frequency channel α of source $x(t)$ and the amplitude time series $|\hat{y}_\beta(t)|$ at frequency channel β of source $y(t)$. Performing this operation for all frequency pairs (α, β), we obtain the amplitude modulation correlation matrix $\mathbf{C}^{xy} \equiv \left[c^{xy}_{\alpha\beta}\right]$ whose elements are defined as $c^{xy}_{\alpha\beta} \equiv c(\hat{x}_\alpha, \hat{y}_\beta)$,

$$\begin{aligned} c(\hat{x}_\alpha, \hat{y}_\beta) &\equiv E\left\{\xi^x_\alpha(t), \xi^y_\beta(t)\right\}, \\ \xi^x_\alpha(t) &\equiv |\hat{x}_\alpha(t)| - E\left\{|\hat{x}_\alpha(t)|\right\}, \\ \xi^y_\beta(t) &\equiv |\hat{y}_\beta(t)| - E\left\{|\hat{y}_\beta(t)|\right\}. \end{aligned}$$

Ideally, this matrix should be identically zero if sources $x(t)$ and $y(t)$ represent two different speech signals. However, one might argue that in the case of two sentences spoken

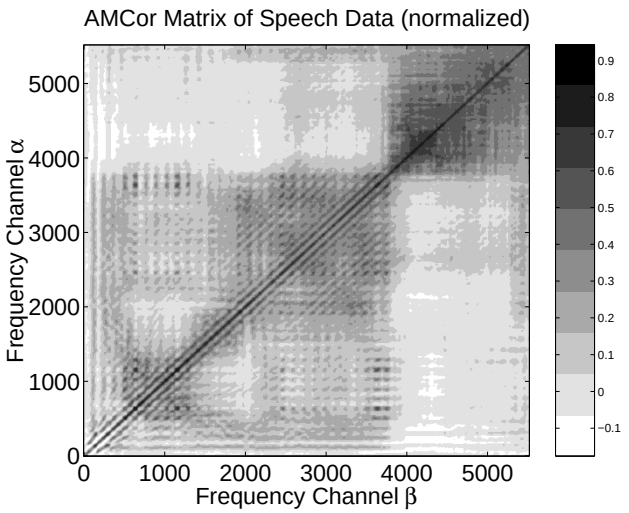

Figure 2: Amplitude modulation correlation ('AMCor') across frequency *within a single* speech signal. AMCor is present even across distant frequencies, e.g., 1000Hz and 3000Hz.

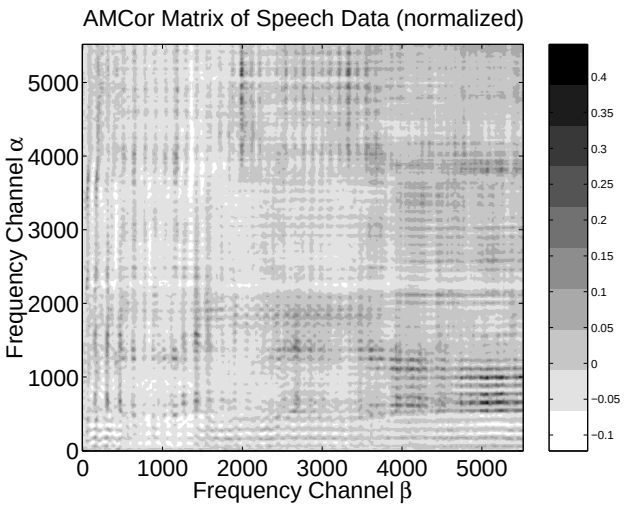

Figure 3: Amplitude modulation correlation ('AMCor') across frequency *between two different* speech signals which were spoken by the same speaker. Comparing the figure to Fig. 2 and noting the different gray–level scales, it is concluded that AMCor between different signals vanishes upto finite size artifacts.

by the same speaker or in the same language, it might not vanish due to speaker– or language–characteristics. Fig. 3 shows an example of an AMCor Matrix for two different sentences spoken by the same speaker in the same language. Clearly, the matrix is close to zero compared to the corresponding AMCor matrix (Fig. 2) computed from the first of the two sentences only. This is also reflected in the ratio of the matrices' squared Frobenius norm which is 24.8.

3. BLIND SOURCE SEPARATION

The convolutive blind source separation problem (Eq. 1) can be recast in the frequency domain using the spectrogram. Provided that the frames for computing the short–time spectra are sufficiently long, Eq. 1 can be approximated by a set of Λ equations, each describing a matrix multiplication in a single frequency band:

$$\hat{\mathbf{x}}_\alpha(t) = \hat{\mathbf{A}}_\alpha \hat{\mathbf{s}}_\alpha(t), \quad \alpha = 1, \ldots, \Lambda.$$

$\hat{\mathbf{x}}_\alpha(t) \equiv [\hat{x}_{1,\alpha}(t), \ldots, \hat{x}_{N,\alpha}(t)]^T$ denotes the vector of the mixed spectrograms at frequency α and time t; $\hat{\mathbf{s}}_\alpha(t) \equiv [\hat{s}_{1,\alpha}(t), \ldots, \hat{s}_{M,\alpha}(t)]^T$ is the corresponding vector for the source signals; the (i,j)–element $\hat{a}_{ij,\alpha}$ of matrix $\hat{\mathbf{A}}_\alpha \equiv [\hat{a}_{ij,\alpha}]$ denotes the value at frequency α of the room transfer function from source j to microphone i. The goal of blind source separation is to find a set of Λ matrices $\hat{\mathbf{W}}_\alpha$, $\alpha = 1, \ldots, \Lambda$ which reconstruct the sources from knowledge of the mixtures only by applying the linear unmixing model to each frequency channel:

$$\hat{\mathbf{u}}_\alpha(t) = \hat{\mathbf{W}}_\alpha \hat{\mathbf{x}}_\alpha(t). \tag{2}$$

The unmixed signals $\hat{\mathbf{u}}_\alpha(t) \equiv [\hat{u}_{1,\alpha}(t), \ldots, \hat{u}_{M,\alpha}(t)]^T$ are required to resemble the original signals upto a permutation and rescaling:

$$\hat{\mathbf{u}}_\alpha(t) = \mathbf{P}_\alpha \mathbf{D}_\alpha \hat{\mathbf{s}}_\alpha(t).$$

The unknown permutation and rescaling is denoted by the permutation matrix $\mathbf{P}_\alpha$ and the diagonal matrix $\mathbf{D}_\alpha$, respectively, both of which can in general depend on frequency α. It must be ensured that the reconstructed signals' ordering with respect to the original signals is the same in every frequency channel, i.e.

$$\mathbf{P} = \mathbf{P}_1 = \mathbf{P}_2 = \ldots = \mathbf{P}_\Lambda.$$

If the unknown permutations $\mathbf{P}_\alpha$ and $\mathbf{P}_\beta$ are different at frequency channels α and β, then the components of the unmixed vectors $\hat{\mathbf{u}}_\alpha(t)$ and $\hat{\mathbf{u}}_\beta(t)$ do not consistently correspond to the components of the source vectors $\hat{\mathbf{s}}_\alpha(t)$ and $\hat{\mathbf{s}}_\beta(t)$. Hence, a time–domain reconstruction of the source signals would be impossible. We term different permutations $\mathbf{P}_\alpha$ and $\mathbf{P}_\beta$ in different frequency channels 'local permutation', in contrast to the 'global permutation' $\mathbf{P}$ which is identical for all frequencies. The unknown global permutation is still present after the local permutations have been solved, however, it does not hinder time–domain reconstruction of the source signals. Hence, the freedom of arbitrary global permutation $\mathbf{P}$ is fixed by assigning the unmixed signal $\hat{u}_{i,\alpha}(t)$ to the i–th source.

The freedom of arbitrary rescaling, introduced by the diagonal matrix $\mathbf{D}_\alpha$, is fixed by modeling the direct paths as unity: $\hat{a}_{ii,\alpha} \equiv 1$. Hence, we reconstruct the i–th unmixed signal as the corresponding source signal's component received by the i–th microphone.

4. AMPLITUDE MODULATION DECORRELATION TECHNIQUE FOR BLIND SOURCE SEPARATION

Amplitude modulation decorrelation can be employed in an elegant manner in order to solve both the blind source

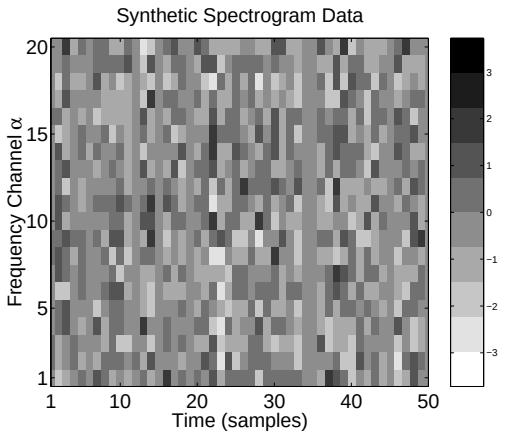

Figure 4: Synthetic source spectrogram containing amplitude modulation correlation across different frequencies. Within each frequency channel, the data has the statistics of a Gaussian i.i.d. random variable.

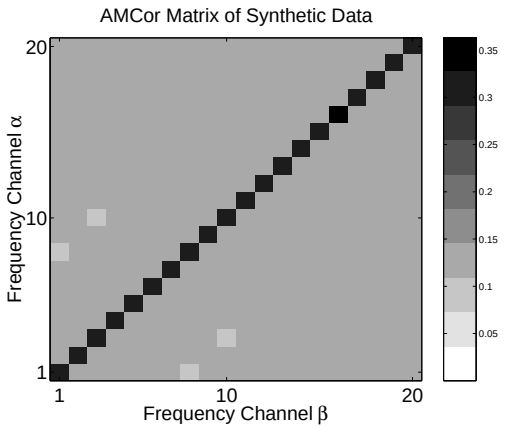

Figure 5: Amplitude modulation correlation ('AMCor') matrix of synthetic source spectrogram. AMCor is non–zero across different frequencies. The peak at the diagonal corresponds to the signal power in each frequency.

separation problem (Eq. 2) and the problem of local permutations simultaneously. Three features of our technique are particularly interesting:

1. Amplitude modulation decorrelation constitutes a distinct criterion for blind source separation. It allows for separation of, e.g., synthetic source signals containing data of Gaussian i.i.d. statistics *within* each frequency channel and amplitude modulation correlation *across* frequencies.

2. Local permutations are penalized by amplitude modulation decorrelation, hence, the problem of local permutations is solved.

3. Since amplitude modulation decorrelation imposes a high number ($O\left(M^2\Lambda^2\right)$) of constraints on the unmixed signals, it achieves a good quality of separation for real–world data.

We now turn to a quantitative description of the amplitude modulation decorrelation technique for blind source separation. In a preprocessing step second order correlations between the microphone signals are removed. It is well–known that decorrelation is not sufficient for source separation (see, e.g., [6]). Rather, additional constraints on the unmixed signals are needed. The amplitude modulation decorrelation principle provides the constraint that the amplitude modulation correlation across any two frequency channels of any two *different* unmixed signals must vanish:

$$c_{\alpha\beta}^{ij} \equiv c\left(\hat{u}_{i,\alpha}, \hat{u}_{j,\beta}\right) = 0 \quad \forall \alpha, \beta, i, j \neq i. \tag{3}$$

Hence, the amplitude modulation correlation matrix $\mathbf{C}^{ij} \equiv \left[c_{\alpha\beta}^{ij}\right]$, $i \neq j$, must be minimized which is done by minimizing its squared Frobenius norm. The cost–function to be minimized with respect to the unmixing matrices $\hat{\mathbf{W}}_\alpha$, $\alpha = 1, \ldots, \Lambda$, is given by

$$H\left(\left\{\hat{\mathbf{W}}_\alpha\right\}\right) = \sum_{i,j\neq i} tr\left(\left[\mathbf{C}^{ij}\right]^T \mathbf{C}^{ij}\right) \tag{4}$$

$$= \sum_{i,j\neq i,\alpha,\beta} c\left(\hat{u}_{i,\alpha}, \hat{u}_{j,\beta}\right)^2. \tag{5}$$

$H\left(\left\{\hat{\mathbf{W}}_\alpha\right\}\right)$ is also referred to as *cumulative amplitude modulation correlation* ('cumulative AMCor').

Clearly, Eq. 3 constitutes a necessary condition for source separation. While we do not give a rigorous prove under which conditions it is also a sufficient condition, experiments performed with synthetic data and speech signals demonstrate that in practice minimization of cumulative AMCor is indeed sufficient in order to achieve a good quality of source separation.

The cost–function (Eq. 4) has the desirable property that it penalizes local permutations. A local permutation between unmixed signals i and j ($j \neq i$) at frequencies α and β manifests in a correlation $c\left(\hat{u}_{i,\alpha}, \hat{u}_{j,\beta}\right)^2 > 0$ which increases the value of $H\left(\left\{\hat{\mathbf{W}}_\alpha\right\}\right)$. Hence, the minimum of $H\left(\left\{\hat{\mathbf{W}}_\alpha\right\}\right)$ corresponds to the optimal solution in which no local permutations occur.

Minimization of Eq. 4 can be performed by gradient based optimization methods. The gradient is given by

$$\delta\hat{\mathbf{W}}_\alpha \propto E\left\{\boldsymbol{\theta}_\alpha(t)\,\hat{\mathbf{x}}_\alpha^H(t)\right\} \tag{6}$$

where

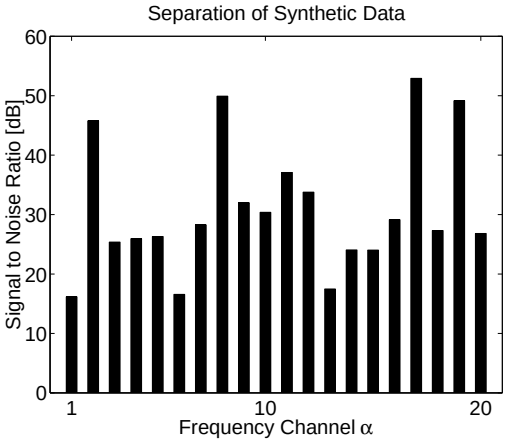

Figure 6: Signal separation of synthetic signals. Local permutations do not occur; they would result in a change of sign of the signal to noise ratio (in dB) at different frequencies.

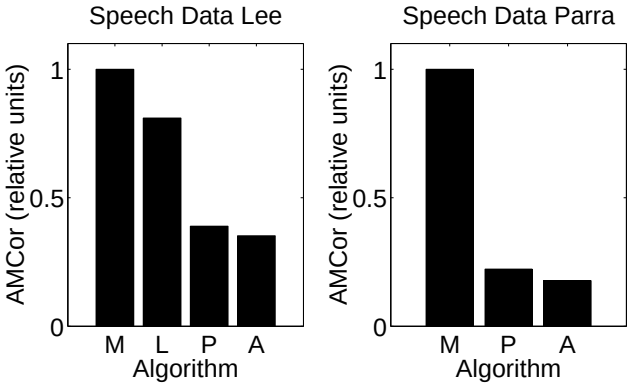

Figure 7: Separation of real–room recordings of speech by different algorithms. Left side refers to speech data recorded by Lee [8], right side to speech data recorded by Parra [11]. The cumulative amplitude modulation correlation ('AMCor') is shown for the amplitude modulation decorrelation algorithm ('A'), Lee's algorithm ('L', [7], results for Lee's speech data only) and Parra's algorithm ('P', [12]), relative to the mixed signals ('M').

$$
\begin{aligned}
\boldsymbol{\theta}_\alpha(t) &\equiv [\theta_{1,\alpha}(t), \ldots, \theta_{M,\alpha}(t)]^T, \\
\theta_{i,\alpha}(t) &\equiv \frac{\hat{u}_{i,\alpha}(t)}{|\hat{u}_{i,\alpha}(t)|} \sum_{i,j\neq i} \sum_{\beta} c(\hat{u}_{i,\alpha}, \hat{u}_{j,\beta})\, \xi_{j,\beta}(t), \\
c(\hat{u}_{i,\alpha}, \hat{u}_{j,\beta}) &\equiv E\left\{\xi_\alpha^{u_i}(t)\, \xi_\beta^{u_j}(t)\right\}, \\
\xi_\alpha^{u_i}(t) &\equiv |\hat{u}_{i,\alpha}(t)| - E\{|\hat{u}_{i,\alpha}(t)|\}, \\
\xi_\beta^{u_j}(t) &\equiv |\hat{u}_{j,\beta}(t)| - E\{|\hat{u}_{j,\beta}(t)|\}.
\end{aligned}
$$

H denotes transposition and complex conjugation.

The 'natural' [1] or 'equivariant' [5] gradient $\tilde{\delta}\hat{\mathbf{W}}_\alpha$ is derived from Eq. 6 by multiplication with $\hat{\mathbf{W}}_\alpha^H \hat{\mathbf{W}}_\alpha$ from the right, resulting in

$$
\tilde{\delta}\hat{\mathbf{W}}_\alpha \propto E\left\{\boldsymbol{\theta}_\alpha(t)\, \hat{\mathbf{u}}_\alpha^H(t)\, \hat{\mathbf{W}}_\alpha\right\}.
$$

We note that Eq. 4 can easily be extended to incorporate time–delayed amplitude modulation decorrelation of the form

$$
c_\tau(\hat{u}_{i,\alpha}, \hat{u}_{j,\beta}) \equiv E\left\{\xi_\alpha^{u_i}(t)\, \xi_\beta^{u_j}(t-\tau)\right\} = 0.
$$

In practice it has proven to be beneficial to minimize Eq. 4 with respect to $\hat{\mathbf{W}}_\alpha$ for fixed frequency α, keeping $\hat{\mathbf{W}}_\beta$ constant for all different frequencies $\beta \neq \alpha$. After optimization of $\hat{\mathbf{W}}_\alpha$ another frequency, α', is selected to be optimized, while then $\hat{\mathbf{W}}_\beta$ is kept constant for all different frequencies $\beta \neq \alpha'$.

5. EXPERIMENTS

5.1. Synthetic data

In this section we construct source spectrograms which contain amplitude modulation correlation *across* different frequency channels. *Within* each frequency channel, their signal statistics is that of a Gaussian independent identically distributed (i.i.d.) random variable. Hence, amplitude modulation correlation is the only cue which can be employed to separate mixtures of the source spectrograms. It is shown that our algorithm successfully separates mixtures of the source signals. Local permutations do not occur.

For each source i, $i = 1, \ldots, M$, we generate Gaussian i.i.d. data $\zeta_{i,\alpha}(t)$ of variance one. Amplitude modulation correlation is introduced by multiplying all $\zeta_{i,\alpha}(t)$ for a particular source i with a common modulator $\mu_i(t)$ which is independent of frequency α. The synthetic source spectrograms are defined as

$$
\hat{s}_{i,\alpha}(t) = \mu_i(t)\, \zeta_{i,\alpha}(t).
$$

Since $\mu_i(t)$ is an i.i.d. random variable of uniform distribution, the $\hat{s}_{i,\alpha}$ have Gaussian probability density function for all i and α. In order to remove finite size artifacts, a non–linear transformation is applied which ensures that the histogram for each $\hat{s}_{i,\alpha}(t)$ has Gaussian shape. Fig. 4 displays a resulting synthetic source spectrogram. The corresponding amplitude modulation correlation matrix is shown in Fig. 5.

Separation of this data–set is possible only by taking into account *across*–frequency information; it is not separable by looking at isolated frequency channels. Since the amplitude modulation decorrelation algorithm exploits modulation information across different frequencies, it successfully separates the data. The accomplished signal separation is displayed in Fig. 6 for a mixture of two source spectra with $\Lambda = 20$ frequencies. Signal separation is measured by the frequency–dependent signal to noise ratio ('SNR') which is defined as the ratio of the desired signals' power and the interfering signals' power contained in the unmixed signal. The mixing matrix $\hat{\mathbf{A}}_\alpha$ was chosen at random and independently for each frequency. From this figure it is also clear that local permutations do not occur, since they would

result in a change of sign of the SNR (in dB) at different frequencies.

5.2. Speech data

We apply the algorithm to two publicly available real–room recordings of speech signals. The quality of separation is evaluated and compared to the quality accomplished by previous algorithms on the same data.

The signals were obtained from Lee [8] and Parra [11]. Since only the mixed signals are known, but not the source signals, it is not possible to compute the signal to noise ratio for the reconstructed signals. Instead, the cumulative amplitude modulation correlation (Eq. 4) is displayed in Fig. 7 for the mixed signals and for separation results from different algorithms. Total energy of the signals was normalized prior to computing the cumulative AMCor. The amplitude modulation decorrelation algorithm (indicated in Fig. 7 by 'A') significantly reduces the level of cumulative AMCor relative to the mixed signals (indicated by 'M'). It performs better than Lee's Algorithm ('L', [7]) and slightly better than Parra's algorithm ('P', [12]). Of course, the value of cumulative AMCor in the unmixed signals is not as good a measure of separation as the signal to noise ratio. However, perceived quality of separation corresponds to the values given in Fig. 7. The signals referred to can be obtained from our web–page [2].

6. CONCLUSION

Based on the notion of amplitude modulation correlation ('AMCor') in speech signals, a novel cost–function for convolutive blind source separation is proposed. The derived amplitude modulation *de*correlation algorithm successfully separates synthetic data and real–room recordings of speech. The algorithm's main features can be summarized as follows: Synthetic spectrograms with data of Gaussian i.i.d. statistics in each frequency channel can be separated if AMCor is present across frequency channels. Local permutations of the unmixed signals are inherently prevented by the algorithm. A large number of constraints on the unmixed signals results in a good separation of echoic speech recordings.

Our discussion has focused on speech signals. However, it is conceivable that various natural signals contain amplitude modulation correlation due to the physical systems that generate them.

7. ACKNOWLEDGMENT

This work was supported by the Deutsche Forschungsgemeinschaft (DFG).

8. REFERENCES

[1] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In D. Touretzky, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 757–763, 1996.

[2] Jörn Anemüller. Ica 2000 demo page. http: //medi.uni-oldenburg.de/ members/ ane/ ica2000.

[3] Jörn Anemüller. Correlated modulation: a criterion for blind source separation. In *Joint meeting of the Acoustical Society of America and the European Acoustics Association*, Berlin, March 1999.

[4] Jörn Anemüller and Tino Gramss. On-line blind separation of moving sound sources. In J. F. Cardoso, Ch. Jutten, and Ph. Loubaton, editors, *ICA '99*, pages 331–334, Aussois, France, January 1999.

[5] Jean-François Cardoso and Beate Hvam Laheld. Equivariant adaptive source separation. *IEEE Transactions on signal processing*, 44:3017–3030, 1996.

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

[7] T.-W. Lee, A. Ziehe, R. Orglmeister, and T. J. Sejnowski. Combining time-delayed decorrelation and ica: towards solving the cocktail party problem. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1249–1252, Seattle, May 1998.

[8] Te-Won Lee. http: //tesla-e0.salk.edu/ ∼tewon/ Blind/ Demos/ rss_mA.wav and rss_mB.wav.

[9] Noboru Murata, Shiro Ikeda, and Andreas Ziehe. An approach to blind source separation based on temporal structure of speech signals. Technical Report 98-2, BSIS, Riken Brain Science Institute, April 1998.

[10] A. V. Oppenheim and R. W. Schaefer. *Digital signal processing*. Prentice-Hall, 1989.

[11] Lucas Parra. http: //www.sarnoff.com/ career_move/ tech_papers/ papers/ tvin1.wav and tvin2.wav.

[12] Lucas Parra and Clay Spence. Convolutive blind source separation based on multiple decorrelation. *IEEE Transaction on Speech and Audio Processing*, 1998, submitted.

[13] Erwin Paulus. *Sprachsignalverarbeitung: Analyse, Erkennung, Synthese*. Spektrum Akademischer Verlag, Heidelberg, Berlin, 1998.

[14] Ch. Servière. Blind source separation in presence of spatially correlated noises. In J. F. Cardoso, Ch. Jutten, and Ph. Loubaton, editors, *ICA '99*, pages 497–502, 1999.