IDIAP RESEARCH REPORT

Adaptive Beamforming with a Maximum Negentropy Criterion

Kenichi Kumatani [a] John McDonough [b]
Dietrich Klakow [b] Philip N. Garner [a]
Weifeng Li [a]

IDIAP–RR 08-06

April 2008

[a] IDIAP Research Institute, Martigny, Switzerland
[b] Spoken Language Systems at Saarland University in Saarbrücken, Germany

IDIAP Research Institute **www.idiap.ch**
Av. des Prés–Beudin 20 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 08-06

Adaptive Beamforming with a Maximum Negentropy Criterion

Kenichi Kumatani John McDonough Dietrich Klakow Philip N. Garner
Weifeng Li

April 2008

Abstract. n this paper, we address an adaptive beamforming application in realistic acoustic conditions. After the position of a speaker is estimated by a speaker tracking system, we construct a subband-domain beamformer in *generalized sidelobe canceller* (GSC) configuration. In contrast to conventional practice, we then optimize the *active weight vectors* of the GSC so as to obtain an output signal with *maximum negentropy* (MN). This implies the beamformer output should be as non-Gaussian as possible. For calculating negentropy, we consider the Γ and the generalized Gaussian (GG) pdfs. After MN beamforming, Zelinski post-filtering is performed to further enhance the speech by removing residual noise. Our beamforming algorithm can suppress noise and reverberation without the signal cancellation problems encountered in the conventional adaptive beamforming algorithms. We demonstrate the effectiveness of our proposed technique through a series of far-field automatic speech recognition experiments on the *Multi-Channel Wall Street Journal Audio Visual Corpus* (MC-WSJ-AV). On the MC-WSJ-AV evaluation data, the delay-and-sum beamformer with post-filtering achieved a word error rate (WER) of 16.5%. MN beamforming with the Γ pdf achieved a 15.8% WER, which was further reduced to 13.2% with the GG pdf, whereas the simple delay-and-sum beamformer provided a WER of 17.8%.

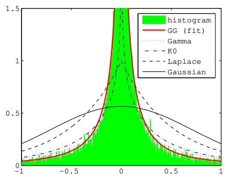

Figure 1: The likelihoods of pdfs and histogram of real parts of subband components.

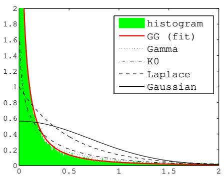

Figure 2: The likelihoods of pdfs and histogram of magnitude in the subband domain.

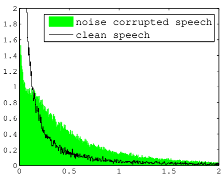

Figure 3: Histogram of subband magnitudes of clean speech and speech corrupted by noise

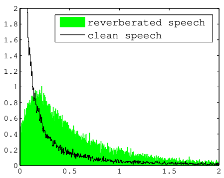

Figure 4: Histograms of subband magnitudes of clean speech and speech corrupted by reverberation.

1 Introduction

There has been great interest in microphone array processing for hands-free automatic speech recognition (ASR) [1, 2]. A conventional beamformer in *generalized sidelobe canceller* (GSC) configuration is structured such that the direct signal from a desired direction is undistorted [3, §14.5]. Subject to this *distortionless constraint*, the total output power of the beamformer is minimized through the adjustment of an *active weight vector*, which effectively places a null on any source of interference, but can also lead to undesirable *signal cancellation* [4]. To avoid the latter, the adaptation of the active weight vector is typically halted whenever the desired source is active.

In this work, we consider negentropy as a criterion for estimating the active weight vectors in a GSC. Negentropy indicates how far a probability density function (pdf) of a particular signal is from Gaussian. In other words, it represents the degree of super-Gaussianity of a pdf [5]. The pdf of speech is in fact super-Gaussian [2, 6], but it becomes closer to Gaussian when the speech is corrupted by noise or reverberation. Hence, in adjusting the active weight vector of the GSC to provide a signal with the highest possible negentropy, it is possible to remove or suppress noise and reverberation. For calculating negentropy, we consider the Γ and the generalized Gaussian (GG) pdfs. After MN beamforming, Zelinski post-filtering is performed to further enhance the speech by removing residual noise [7]. We demonstrate the effectiveness of our proposed technique through a series of far-field ASR experiments on the *Multi-Channel Wall Street Journal Audio Visual Corpus* (MC-WSJ-AV) [1], a corpus of multi-modal data captured with real far-field sensors, in a realistic acoustic environment, and with real speakers.

The balance of this work is organized as follows. Section 2 describes the characteristics of super-Gaussian pdfs and illustrates that subband samples of clean speech are super-Gaussian distributed. Speech corrupted with noise or reverberation, however, becomes more nearly Gaussian distributed. Section 3 reviews the definition of the negentropy. In Section 4, we discuss our maximum negentropy beamforming criterion and derive the objective functions for estimating the active weight vectors. Thereafter we comment on the fact that the proposed algorithm, unlike conventional beamformers, does not suffer from the signal cancellation problem. In Section 5, we present the results of far-field automatic speech recognition experiments. Finally, in Section 6, we present our conclusions and plans for future work.

2 Modeling Subband Samples of Speech with Super-Gaussian Probability Density Functions

The fact that the pdf of speech is super-Gaussian has often been reported in the literature [2, 6]. Noise, on the other hand, is more nearly Gaussian-distributed. The pdf of the sum of even two super-Gaussian random variables will be more nearly Gaussian than either of the two original sources. Based on this observation, we hope to remove interference signals and extract a target speech signal by making the pdf of the beamformer's output as super-Gaussian as possible.

A plot of the likelihood of the Gaussian and four super-Gaussian univariate pdfs is provided in Fig. 1, where the parameters of the GG pdfs are estimated from the actual speech data. From the figure, it is clear that the Laplace, K_0, Γ and GG pdfs, exhibit the "spikey" and "heavy-tailed" characteristics that are typical of super-Gaussian pdfs. Fig. 1 also shows the histogram of the real parts of subband speech samples. Fig. 2 shows the histogram of magnitude in the subband domain with a plot of the likelihood of the pdfs. In both figures, the clean speech recorded with the close-talking microphone (CTM) in the Speech Separation Challenge, Part II (SSC2) development set [1] was used for the histograms, and the parameters of the GG pdf were estimated from this development data. It is clear from Fig. 1 and 2 that the distribution of clean speech is not Gaussian but super-Gaussian. The figures also suggest that the GG pdf can be suitable for modeling speech.

Fig. 3 shows histograms of magnitude of clean speech and noise corrupted speech in the subband domain. It is clear from this figure that the pdf of the noise corrupted speech has less probability mass around the center spike, but more probability mass in intermediate regions. This indicates that the pdf of the noise-corrupted signal, which is in fact the sum of the speech and noise signals, is closer to Gaussian than that of clean speech. Fig. 4 shows histograms of magnitude of clean speech and reverberated speech. We can observe from Fig. 4 that the pdf of reverberated speech is also closer to Gaussian than the original clean speech. Interestingly, Fig. 4 shows that the peak of the histogram of the speech is shifted from zero to the right by the effects of reverberation. These observations support the hypothesis that performing acoustic beamforming to obtain an enhanced speech signal that is maximally non-Gaussian is an effective way to suppress the distortions introduced by noise and reverberation.

2.1 Super-Gaussian pdf derived from the Meijer G-function

As noted by Brehm and Stammler [8], it is useful to model speech as a *spherically-invariant random process* (SIRP), because such processes are completely characterized by their first and second order moments. Moreover, Brehm and Stammler [8] noted that the Laplace, K_0, and Γ pdfs can all be represented as *Meijer G-functions*, which is useful for two reasons. Firstly, this implies that multivariate pdfs of all orders can be readily derived from the univariate pdf. Secondly, such variates can be extended to the case of complex r.v.s, which is essential for our current development. Of the three pdfs derived from the Meijer G-function, we chose to use the Γ pdf for the experiments reported in Section 5, as it achieved the highest likelihood on the development data [2]. For the Γ pdf, the complex univariate pdf *cannot* be expressed in closed form in terms of elementary or even special functions. As explained in [2], however, it is possible to derive Taylor series expansions that enable the required variates to be calculated to arbitrary accuracy.

2.2 Generalized Gaussian pdf

Due to its definition as a contour integral, finding maximum likelihood estimates for the parameters of a Meijer G-function must necessarily devolve to a grid search over the parameter space [8]. Instead, it may be better to use a simple super-Gaussian pdf whose parameters can easily be adjust so as to match the pdf of actual speech. The generalized Gaussian (GG) pdf is well-known and finds frequent application in the fields of *blind source separation* (BSS) and *independent component analysis* (ICA). The GG pdf with zero mean for a real-valued r.v. y is by definition

$$p_{\mathrm{GG}}(y) \triangleq \frac{1}{2\Gamma(1+1/p)A(p,\hat{\sigma})} \exp\left[-\left|\frac{y}{A(p,\hat{\sigma})}\right|^p\right], \tag{1}$$

$$\text{where} \qquad A(p,\hat{\sigma}) = \hat{\sigma}\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{1/2}, \tag{2}$$

$\Gamma(.)$ is the gamma function and p is the shape parameter, which controls how fast the tail of the pdf decays. Note that the GG with $p = 1$ corresponds to the Laplace pdf, and that setting $p = 2$ yields the conventional Gaussian pdf, whereas in the case of $p \rightarrow +\infty$ the GG pdf converges to a uniform distribution.

The differential entropy of the GG pdf for the real-valued r.v. y is obtained with the help of *Mathematica* [9] as

$$H_{\mathrm{GG}}(y) \quad = \quad 1/p + \log\left[2\Gamma(1+1/p)A(p,\hat{\sigma})\right]. \tag{3}$$

Among several methods for estimating the shape parameter p of the GG pdf [10], the moment and maximum likelihood (ML) methods are arguably the most straightforward. In this work, we use the the moment method in order to initialize the parameters of the GG pdf and then update them with the ML estimate [10]. The shape parameters are estimated from training samples offline and are held fixed during the adaptation of the active weight vector. The shape parameters for each subband are estimated independently, as the optimal pdf is frequency-dependent.

3 Negentropy and Kurtosis

The *entropy* for a continuous complex-valued r.v. Y, which is often called the differential entropy, is defined as

$$H(Y) \triangleq -\int p_Y(v)\log p_Y(v)dv = -\mathcal{E}\left\{\log p_Y(v)\right\}, \tag{4}$$

where $p_Y(.)$ is the pdf of Y. The entropy of a r.v. indicates how much information the observation of the variable provides. Negentropy J for a complex-valued r.v. Y is defined as

$$J(Y) = H(Y_{\mathrm{gauss}}) - H(Y) \tag{5}$$

where Y_{gauss} is a Gaussian variable which has the same variance σ_Y^2 as Y. The entropy of Y_{gauss} can be expressed as

$$H(Y_{\mathrm{gauss}}) = \log\left|\sigma_Y^2\right| + n\left(1+\log 2\pi\right) \tag{6}$$

where n is the dimension of Y. It is well-known that a Gaussian variable has the largest entropy among all r.v.s of equal variance [11, Thm. 7.4.1]. Due to this fact, negentropy is always non-negative and zero only if Y is Gaussian. Hence, negentropy is useful as a measure of non-Gaussianity. Both negentropy and kurtosis are frequently used in the field of ICA as such measures of deviation from Gaussianity. Hyvärinen and Oja [5] noted that negentropy was generally more robust in the presence of outliers than kurtosis. Hence, we adopt negentropy as our measure of choice.

4 Beamforming and Post-Filtering

Consider a subband beamformer in the GSC configuration [3, §14.5] with a post-filter. The output of a beamformer for a given subband can be expressed as

$$Y = \left(\mathbf{w}_{\mathrm{q}} - \mathbf{B}\mathbf{w}_{\mathrm{a}}\right)^H \mathbf{X}, \tag{7}$$

where $\mathbf{w}_{\mathrm{q}}$ is the *quiescent weight vector*, $\mathbf{B}$ is the *blocking matrix*, $\mathbf{w}_{\mathrm{a}}$ is the *active weight vector*, and $\mathbf{X}$ is the input subband *snapshot vector*.

In keeping with the GSC formalism, $\mathbf{w}_{\mathrm{q}}$ is chosen to give unity gain in the *look direction* [3, §14.5]; i.e., to satisfy a *distortionless constraint*. The blocking matrix $\mathbf{B}$ is chosen to be orthogonal to $\mathbf{w}_{\mathrm{q}}$, such that $\mathbf{B}^H\,\mathbf{w}_{\mathrm{q}} = \mathbf{0}$. This orthogonality implies that the distortionless constraint will be satisfied for any choice of $\mathbf{w}_{\mathrm{a}}$. While the active weight vector $\mathbf{w}_{\mathrm{a}}$ is typically chosen to maximize the signal-to-noise ratio (SNR), here we will develop an optimization procedure to find that $\mathbf{w}_{\mathrm{a}}$ *maximizing* the negentropy $J(Y)$ in (5).

In order to calculate negentropy, the variance of the output Y is needed. Substituting (7) into the definition $\sigma_Y^2 = \mathcal{E}\left\{Y\,Y^*\right\}$ of variance, we find

$$\sigma_Y^2 \quad = \quad (\mathbf{w}_\mathrm{q} - \mathbf{B}\mathbf{w}_\mathrm{a})^H \, \boldsymbol{\Sigma}_\mathbf{X} \, (\mathbf{w}_\mathrm{q} - \mathbf{B}\mathbf{w}_\mathrm{a}) \,, \tag{8}$$

where $\boldsymbol{\Sigma}_\mathbf{X}$ is the covariance matrix of $\mathbf{X}$. Maximizing the negentropy criterion yields a weight vector $\mathbf{w}_\mathrm{a}$ capable of canceling interferences including incoherent noise that leaks through the sidelobes without the signal cancellation problems encountered in conventional beamforming. With the weight of the Zelinski post-filter w_z, the final output of the beamformer and post-filter combination is

$$Y_\mathrm{f} = w_z Y = w_z \, (\mathbf{w}_\mathrm{q} - \mathbf{B}\mathbf{w}_\mathrm{a})^H \; \mathbf{X}, \tag{9}$$

where w_z is the frequency-dependent of the post-filter [7].

For the experiments described in Section 5, subband analysis and synthesis were performed with uniform DFT filter banks. For each of the analysis and synthesis banks, a single filter prototype was designed and subsequently modulated in order to minimize each subband aliasing component individually [12].

In conventional beamforming, a *regularization* term is often applied that penalizes large active weights, and thereby improves robustness by inhibiting the formation of excessively large sidelobes [3, §14.6]. Such a regularization term can be applied in the present instance by defining the modified optimization criterion

$$\mathcal{J}(Y;\alpha) = J(Y) + \alpha \|\mathbf{w}_\mathrm{a}\|^2 \tag{10}$$

for some real $\alpha > 0$. For the experiments described in this work, we set $\alpha = 0.01$.

4.1 Estimation of Active Weights under the Γ pdf

Here we describe necessary formulae for estimating the active weight vectors in the case that the Γ pdf assumption is used. In this case, the differential entropy (4) cannot be expressed in closed form. We must, therefore, replace the exact differential entropy with the *empirical differential entropy*

$$H(Y) = -\mathcal{E}\left\{\log p_Y(Y)\right\} \approx -\frac{1}{T}\sum_{t=0}^{T-1} \log p_Y(Y_t). \tag{11}$$

Substituting (11) and (6) into (5), we can express the negentropy as

$$J(Y) = \log\left|\sigma_Y^2\right| + n\,(1 + \log 2\pi) + \frac{1}{T}\sum_{t=0}^{T-1} \log p_Y(Y_t). \tag{12}$$

We maximize the objective function which is the sum of the negentropy and the regularization term. In the absence of a closed-form solution for the $\mathbf{w}_\mathrm{a}$ maximizing the negentropy (12), we must use a numerical optimization algorithm. Such an optimization algorithm typically requires gradient information.

By substituting (12) into (10) and taking the partial derivative on both sides, we obtain the gradient,

$$\frac{\partial \mathcal{J}(Y;\alpha)}{\partial {\mathbf{w}_\mathrm{a}}^*} = \frac{1}{|\sigma_Y^2|}\frac{\partial |\sigma_Y^2|}{\partial {\mathbf{w}_\mathrm{a}}^*} + \frac{1}{T}\sum_{t=0}^{T-1} \frac{1}{p_Y(Y_t)}\frac{\partial p_Y(Y_t)}{\partial {\mathbf{w}_\mathrm{a}}^*} + \alpha \mathbf{w}_\mathrm{a},$$

which is sufficient to implement a numerical optimization algorithm based, for example, on the method of *conjugate gradients* [13, §1.6].

4.2 Estimation of Active Weights under the GG pdf

Unlike the pdfs that can be expressed as Meijer G-functions, the GG pdf cannot be readily extended from the univariate to the multi-variate. Hence, we use the magnitude of beamformer's output as the r.v. for calculating the entropy. By substituting (3) and (6) into (5), we obtain the negentropy

$$J(Y) = \log\left|\sigma_Y^2\right| + n\,(1 + \log 2\pi) - H_{\mathrm{GG}}(|Y|). \tag{13}$$

Similarly, by adding the regularization term and taking partial derivatives on both sides of (13), we can obtain the objective function

$$\frac{\partial \mathcal{J}(Y;\alpha)}{\partial \mathbf{w}_{\mathrm{a}}{}^*} = \frac{1}{\sigma_Y^2}\frac{\partial \sigma_Y^2}{\partial \mathbf{w}_{\mathrm{a}}{}^*} - \frac{1}{\hat{\sigma}_{|Y|}}\frac{\partial \hat{\sigma}_{|Y|}}{\partial \mathbf{w}_{\mathrm{a}}{}^*} + \alpha \mathbf{w}_{\mathrm{a}}, \tag{14}$$

$$\frac{\partial \hat{\sigma}_{|Y|}}{\partial \mathbf{w}_{\mathrm{a}}{}^*} = \frac{p}{T}\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{1}{2}}\left[\frac{p}{T}\sum_{t=0}^{T-1}|Y_t|^p\right]^{\frac{1}{p}-1}\left[\sum_{t=0}^{T-1}|Y_t|^{p-1}\frac{\partial |Y_t|}{\partial \mathbf{w}_{\mathrm{a}}{}^*}\right],$$

$$\text{and} \qquad \frac{\partial |Y_t|}{\partial \mathbf{w}_{\mathrm{a}}{}^*} = -\frac{1}{2|Y_t|}\mathbf{B}^H \mathbf{X} Y_t^*.$$

We can implement a numerical optimization algorithm with the equations described above.

4.3 Discussion about the Signal Cancellation Problem

Conventional adaptive beamforming algorithms determine the optimum weight vector that minimizes the variance of the beamformer's output subject to the distortionless constraint. Such a conventional beamformer would attempt to null out any interfering signal. This, however, can lead to signal cancellation [4] in the case that there is an interference signal which is correlated with the desired signal. In realistic environments, interference signals are highly correlated with a target signal since the target signal is reflected from hard surfaces such as walls and tables. In contrast to conventional beamformers, the MN beamforming algorithm would attempt not only to eliminate interference signals but also strengthen those reflections from the desired source, assuming a sound source is statistically independent of the other sources. Of course, any reflected signal would be delayed with respect to the direct path signal. Such a delay would, however, manifest itself as a phase shift in the subband domain, and could thus be removed through a suitable choice of $\mathbf{w}_{\mathrm{a}}$. Hence, the MN beamformer offers the possibility of steering both nulls and sidelobes; the former towards the undesired signal and its reflections, the latter towards reflections of the desired signal.

5 Experiments

We performed far-field ASR experiments on the MC-WSJ-AV; see [1] for a description of the data collection apparatus. In the single speaker stationary scenario of the MC-WSJ-AV, a speaker was asked to sit or stand in front of a presentation screen and read sentences from different positions. The far-field speech data was recorded with two circular, eight-channel microphone arrays in a reverberant room. In addition to the reverberation, some recordings include significant amounts of background noise. Our test data set for the experiments contains recordings of 10 speakers where each speaker reads approximately 40 sentences taken from the 5,000 word vocabulary WSJ task. This provided a total of 352 utterances which correspond to approximately 43.9 minutes of speech. There are a total of 11,598 word tokens in the reference transcriptions. Prior to beamforming, we first estimated speaker's position with the Orion source tracking system [2, 14]. Based on the average speaker position estimated for each utterance, utterance-dependent active weight vectors $\mathbf{w}_a$ were estimated for the source. The active weight vectors for each subband were initialized to zero for estimation. Iterations of the conjugate gradients algorithm were run on the entire utterance until convergence was achieved. As mentioned previously, Zelinski post-filtering [7] was performed after beamforming. We did four decoding passes on the waveforms obtained with the beamforming algorithms described above. Each pass of decoding used a different acoustic model or speaker adaptation scheme. Speaker adaptation parameters were estimated using the word lattices generated during the prior pass. The details of the speech recognition engine are presented in [14]. Table 1 shows the word error rates (WERs) for every beamforming algorithm. As references, WERs in recognition experiments on speech data recorded with the single distant microphone (SDM) and with the CTM are also given in Table 1. It is clear from Table 1 that every MN beamforming algorithm provides better recognition performance than the simple delay-and-sum beamformer both without (D&S BF) and with Zelinski post-filtering (D&S BF with PF). It is also clear from Table 1 that MN beamforming with the GG pdf assumption (MN BF

Table 1: Word error rates for each beamforming algorithm after every decoding pass.

Beamforming Algorithm	Pass (%WER)			
	1	2	3	4
D&S BF	80.1	39.9	21.5	17.8
D&S BF with PF	79.0	38.1	20.2	16.5
MN BF with Gamma pdf	75.6	34.9	19.8	15.8
MN BF with GG pdf	75.1	32.7	16.5	13.2
SDM	87.0	57.1	32.8	28.0
CTM	52.9	21.5	9.8	6.7

with GG pdf) achieves the best recognition performance. Table 1 suggests that the Γ pdf assumption (MN BF with Γ pdf) has better noise suppression performance than the D&S beamformers. The performance of the negentropy beamformer under a Γ pdf is not as good as that under the GG with frequency-dependent shape factors, inasmuch as this frequency-dependence enables much more accurate modeling of the speech spectra.

6 Conclusions and Future Work

In this work, we have proposed a novel beamforming algorithm based on maximizing negentropy and demonstrated that the proposed method with the GG pdf assumption provided the best recognition performance. In future, we plan to develop an on-line version of the beamforming algorithm presented here.

References

[1] M. Lincoln, I. McCowan, I. Vepa, and H. K. Maganti, "The multi-channel Wall Street Journal audio visual corpus (mc-wsj-av): Specification and initial experiments," in *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Cancun, Mexico, 2005, pp. 357–362.

[2] Kenichi Kumatani, Tobias Gehrig, Uwe Mayer, Emilian Stoimenov, John McDonough, and Matthias Wölfel, "Adaptive beamforming with a minimum mutual information criterion," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 2527–2541, 2007.

[3] Matthias Wölfel and John McDonough, *Distant Speech Recognition*, Wiley, New York, 2008.

[4] Bernard Widrow, Kenneth M. Duvall, Richard P. Gooch, and William C. Newman, "Signal cancellation phenomena in adaptive antennas: Causes and cures," *IEEE Transactions on Antennas and Propagation*, vol. AP-30, pp. 469–478, 1982.

[5] Aapo Hyvärinen and Erkki Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411–430, 2000.

[6] Jan S. Erkelens, Richard C. Hendriks, Richard Heusdens, and Jesp er Jensen, "Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 1741–1752, 2007.

[7] Claude Marro, Yannick Mahieux, and K. Uwe Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Transactions on Speech and Audio Processing*, vol. 6, pp. 240–259, 1998.

[8] Helmut Brehm and Walter Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing*, vol. 12, pp. 119–141, 1987.

[9] Stephen Wolfram, *The Mathematica Book*, Cambridge University Press, Cambridge, 3 edition, 1996.

[10] Mahesh K. Varanasi and Behnaam Aazhang, "Parametric generalized gaussian density estimation," *J. Acoust. Soc. Am.*, vol. 86, pp. 1404–1415, 1989.

[11] Robert G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.

[12] Kenichi Kumatani, John McDonough, Stefan Schacht, Dietrich Klakow, Philip N. Garner, and Weifeng Li, "Filter bank design based on minimization of individual aliasing terms for minimum mutual information subband adaptive beamforming," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, Nevada, U.S.A, 2008.

[13] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.

[14] John McDonough, Kenichi Kumatani, Tobias Gehrig, Emilian Stoimenov, Uwe Mayer, Stefan Schacht, Matthias Woelfel, and Dietrich Klakow, "To separate speech! a system for recognizing simultaneous speech," in *Proc. 4th Joint Workshop on Machine Learning and Multimodal Interaction (MLMI)*, Brno, Czech Republic, 2007.