ADAPTIVE LISTENING ROOM COMPENSATION FOR SPATIAL AUDIO SYSTEMS

S.Spors, H.Buchner and R.Rabenstein

Telecommunications Laboratory
University of Erlangen-Nuremberg
Cauerstrasse 7, 91058 Erlangen, Germany
E-mail: {spors, buchner, rabe}@LNT.de

ABSTRACT

Various kinds of spatial audio systems have been developed that go beyond the current 5.1 standard. Although the approaches are different, they are capable of delivering acoustic immersion to a large number of listeners simultaneously without restricting their activities by headphones, trackers, or alike. In particular, the sweet spot limitation well known from current consumer market systems is repealed. However, these attractive properties could only be realized in listening rooms without any noticeable reverberation. Otherwise, measures have to be taken for the digital compensation of wall reflections. This contribution presents an adaptive compensation technique which is based on the decomposition of sound fields into plane waves. Various advantages are discussed, e.g. improved multichannel adaptation by decoupling of the compensation filters. Finally, results for a specific spatial reproduction method (wave field synthesis) are presented.

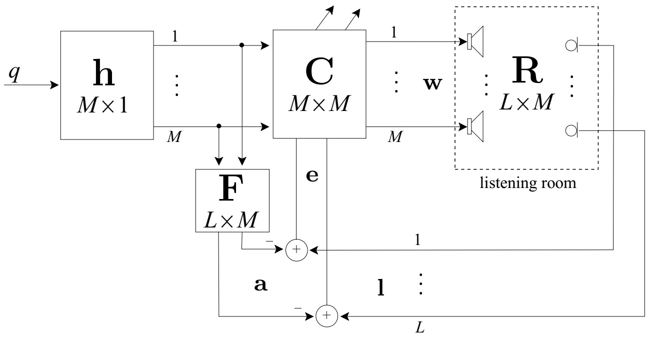

Figure 1: Block diagram of traditional approaches to room compensation

1. INTRODUCTION

1.1 Spatial Audio Systems

Advanced systems for the reproduction of multimedia content or for immersion into virtual realities require spatial audio capabilities that go beyond the current 5.1 standard. These requirements can be summarized very briefly as: Reproduce (i) a number of static or moving sound sources (ii) to an arbitrary number of listeners without restricting their movement and activities (iii) in arbitrary listening rooms.

Requirements (i) and (ii) can be met by spatial audio reproduction schemes which overcome the sweet spot well known from two-channel stereo or 5-channel surround sound systems. Such schemes include advanced panning techniques, Ambisonics, and wave field synthesis. However, the theory behind all these methods always assumes an anechoic listening (or reproduction) room which does not exhibit any reflections. This idealistic assumption conflicts with requirement (iii) calling for arbitrary listing rooms.

A first countermeasure against room reflections is the application of acoustic insulation materials, called passive room compensation in this context. However, it is well known that acoustic insulation gets impractical and costly above an even rather modest level of sound absorption. This holds especially for low frequencies. Although indispensable as a first measure, passive room compensation alone cannot provide a sufficient suppression of listening room reflections for a successful application of spatial audio systems.

A technically more demanding approach applies active acoustical instrumentation (microphones, loudspeakers, amplifiers, etc.) to counteract undesired room reflections (active room compensation). While the amount of microphones and loudspeakers for a successful operation may be quite large, there are synergies to be exploited, since spatial audio systems employ multichannel audio reproduction anyway.

1.2 Traditional active room compensation

Traditional approaches to active room compensation are based on point-wise measurements of the reproduced sound field and the derivation of appropriate compensation filters from these measurements. Fig. 1 displays a generic block diagram. For simplicity, only a single virtual sound source q is shown. For a description of this setup, consider at first the uncompensated case where the matrix of compensation filters $\mathbf{C}$ equals the identity matrix. The vector $\mathbf{h}$ contains the appropriate impulse responses for the generation of the loudspeaker signals $\mathbf{w}$ from the source signal q. The set of filters $\mathbf{h}$ is defined by the specific spatial audio rendering method at hand.

The response of the listening room to the loudspeaker signals can be measured at arbitrary microphone positions. It is described by the matrix $\mathbf{R}$ of room impulse responses which contain also the reflections in the listening room. Placing the microphones at the possible listener positions will give a realistic reproduction of the listener's impression. Due to the room reflections, it may deviate considerably from the intended sound rendering.

This deviation can be quantified by comparing the microphone signals with an idealized matrix of room impulse responses $\mathbf{F}$ for the reflection free case. $\mathbf{F}$ can be analytically derived from the acoustic wave propagation for the free field case. In simple cases, $\mathbf{F}$ models only delays and attenuations. The error signal $\mathbf{e}$ describes the deviation of the auralized wave field from the reflection free case.

Finally, the error $\mathbf{e}$ can be used to control a set of compensation filters $\mathbf{C}$. Applying well-known adaptive filtering methods for error-minimization generates compensated loudspeaker signals $\mathbf{w}$ which – together with the unavoidable room reflections – produces the desired sound impression for the listeners.

The specific approaches for active room compensation vary in the specific loudspeaker and microphone setup, the determination of the spatial filters $\mathbf{h}$ and in the adaptive algorithm for the compensation filters $\mathbf{C}$. However, practical experience has unveiled a number of drawbacks common to all of these approaches: (1) The compensation of room reflections works well only at the microphone positions. This is a result of the limited control and analysis capabilities of typical active room compensation systems. (2) Consequently, the compensation for arbitrary wave fields requires a high number of playback and microphone channels to have a sufficient amount of control. (3) Finally, the adaptation of the compensation filters is not feasible for a system with many playback and reference channels

due to the high dimensionality of the required filter matrix $\mathbf{C}$.
The dimensionality of the filter matrix $\mathbf{C}$ could be lowered by incorporating the auralization operator $\mathbf{h}$ into the compensation filters $\mathbf{C}$ as shown in [1]. But this approach requires one set of compensation filters for each virtual source position, making the synthesis of moving sound sources quite complex.

1.3 Requirements for improved room compensation

From this discussion of traditional active room compensation, a number of requirements for improved methods can be derived. Advanced room compensation methods should

1. derive error signals from the entire listening area, not only from selected points,
2. work with decoupled compensation filters to lower the computational complexity and improve adaptation,
3. be based on a spatial audio system which provides control over the acoustic wave field inside the entire listening area.

We will first concentrate on the first two points in the following section. Then an example for a specific system will be presented in the results.

2. LISTENING ROOM ANALYSIS

Room compensation can also be understood as inverse identification problem. By identifying the room characteristics and performing pre-filtering of the loudspeaker signals according to the inverse room characteristics active room compensation is realized. However, this requires a thorough analysis of the listening room influence inside the entire listening area. The following section will introduce the necessary tools to analyze the listening area in our context. We limit our discussion to two-dimensional wave fields in this paper, but the same principles also apply to three-dimensional wave fields.
Various analysis techniques for acoustic wave fields in reflective media are well studied in the context of seismic imaging. The basic goal of seismic imaging is to derive the position and reflectivity characteristics of layered media in order to identify media boundaries and thus the structure of the analyzed volume. The same principles can also be applied to analyze acoustic wave fields in enclosures as shown in the following.

2.1 Plane wave decomposition

Let us first assume that we have access to the two-dimensional wave field $P(x,y,\omega)$ inside the entire listening area. Discarding the effects of spatial sampling this could be realized by placing microphones inside the entire listening area. Performing a two-dimensional spatial Fourier transform on the pressure field $P(x,y,\omega)$ yields

$$\tilde{P}(k_x,k_y,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y,\omega)\, e^{j(k_x x + k_y y)}\, dx\, dy \qquad (1)$$

where x and y denote the position in Cartesian coordinates, ω denotes the temporal frequency, k_x and k_y the spatial wave numbers. We want to decompose the wave field into its plane wave components with respect to an arbitrary chosen origin. Equation (1) involves all possible wave numbers k_x and k_y and not only those which belong to a plane wave. A plane wave dispersion relation is introduced through a change of the Cartesian coordinate system to a cylindrical coordinate system as follows [2]

$$x = r\cos\alpha \qquad k_x = k\cos\theta \qquad (2)$$

$$y = r\sin\alpha \qquad k_y = k\sin\theta \qquad (3)$$

where r and α denote the position in cylindrical coordinates with respect to the origin, k and θ the wave number and incidence angle of the plane waves. Substitution into equation (1) yields

$$\tilde{P}(\theta,\omega) = \mathscr{P}\{P(r,\alpha,\omega)\} = \int_0^{\infty}\int_0^{2\pi} P(r,\alpha,\omega) e^{jkr\cos(\theta-\alpha)} r\, d\alpha\, dr \qquad (4)$$

which decomposes the recorded wave field into its plane wave components. This transformation is termed as *plane wave decomposition* and is well known from seismic imaging. One benefit of this approach is that this signal representation is independent from the particular geometry used for analysis.
It can be shown that the plane wave decomposition is equivalent to a Radon transformation. In the case of a radially symmetric wave field $P(r,\omega)$ with respect to the origin it equals the Fourier-Bessel (Hankel) transformation [2].
Plane wave components can be easily extrapolated to other positions. This operation is often termed as *plane wave extrapolation* or *wave field extrapolation* [3]. The acoustic pressure at an arbitrary point can the be calculated by summing up all extrapolated components as follows

$$P(r,\alpha,\omega) = \mathscr{P}^{-1}\{\tilde{P}(\theta,\omega)\} = \int_0^{2\pi} \tilde{P}(\theta',\omega) e^{-jkr\cos(\alpha-\theta')} d\theta', \qquad (5)$$

This allows, in principle, to extrapolate a recorded wave field to arbitrary points without loss of information. Wave field extrapolation can be used to extrapolate a measured field to the loudspeaker positions for reproduction purposes or to create a complete image of the captured sound field.
However, recording $P(x,y,\omega)$ for the entire listening area with microphones is not feasible in our context. On the one hand this would require a quite high number of microphones, on the other hand the microphones would occupy the listening positions. A solution to this problem is provided by the Kirchoff-Helmholtz (KH) integral [3]. The KH integral states that at any listening point within a source-free volume V the sound pressure can be calculated if both the sound pressure and its gradient are known on the surface S enclosing the volume. The original formulation of the Kirchoff-Helmholtz integral is given for three-dimensional volumes V, but it can also be formulated for the two-dimensional case [3]. Measurement of the acoustic pressure and its gradient on a closed contour around the listening area are sufficient to describe the entire area. However the degeneration to two-dimensions has drawbacks: amplitude errors and interference of elevated sources into the plane wave components. The first effect is caused by amplitude difference between a 2D and a 3D point source. The second effect is related to the 2D geometry of the analysis array. The fact that microphones can only be mounted at discrete positions can result in spatial aliasing due to spatial sampling. Implementations of the plane wave decomposition for different microphone array geometries can be found in [4].

2.2 Listening room reflections

To achieve the desired decoupling of the compensation filters, a quantitative description of the listening room influence inside the listening area is required. The plane wave concept introduced in the previous section provides an efficient tool for this purpose. Its application requires a mild assumption which holds quite well for linear room acoustics. Consider an excitation of the listening room by a plane wave with a certain incidence angle. Now assume that the reflected wave field resulting from this excitation is a superposition of delayed and weighted plane waves from all possible directions. The reflected wave field $\tilde{P}_{refl}(\theta,\omega)$ for one particular plane wave component can then be expressed as a convolution between a plane wave room transfer function $\tilde{r}(\theta,\omega)$ and the incident plane wave $\tilde{P}_{inc}(\theta,\omega)$ as follows

$$\tilde{P}_{refl}(\theta,\omega) = \sum_{\theta'} \tilde{r}(\theta',\omega)\, \tilde{P}_{inc}(\theta'-\theta,\omega) \qquad (6)$$

The overall response of the room results from a multichannel convolution in the plane wave domain.

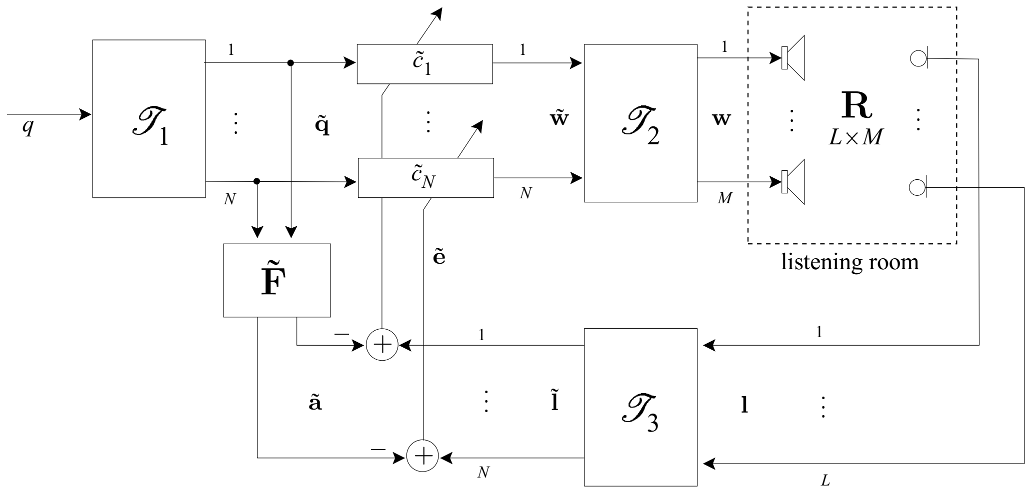

Figure 2: Block diagram of proposed approach

2.3 Decoupling

The plane wave decomposition in combination with the KH relation provides an efficient way to fulfill the first requirement from section 1.3. Until now we still have not addressed the second requirement. As stated we need a transformation that decouples the room transfer matrix $\mathbf{R}$ and thus the compensation filters. Applying a Fourier transform with respect to the angular parameter θ of the plane wave decomposition

$$\tilde{P}(k_\theta, \omega) = \mathscr{F}_\theta \left\{ \tilde{P}(\theta, \omega) \right\} \tag{7}$$

where k_θ denotes the angular wave number, provides the decoupled signal components as desired.

3. A NOVEL APPROACH TO ROOM COMPENSATION

Figure 2 shows a block diagram of our proposed approach. The basic idea is to introduce a set of spatial transformations $\mathscr{T}_1$ to $\mathscr{T}_3$ that orthogonalize the listening room response $\mathbf{R}$. As a consequence, the matrix of compensation filters $\mathbf{C}$ is decomposed into a set of compensation filters $\tilde{c}_i$, each acting on only one spatial signal component. The adaption of these compensation filters is then performed independently for each spatially transformed component. The number of compensation filters that have to be adapted is lowered significantly compared to the traditional approaches described in section 1.2. Thus the complexity of the filter adaption is reduced.
In the following we will specify the transformations $\mathscr{T}_1$ to $\mathscr{T}_3$. The transformation $\mathscr{T}_1$ transforms the virtual source q to be auralized into the frequency components of the plane wave domain. Suitable spatial source models, like point source or plane wave propagation, allow a closed-form solution of this equation [2]. Using the generic plane wave decomposition (4) it is also possible to prescribe complex wave fields as desired source signal. The transformed signals are then pre-filtered by the room compensation filters $\tilde{c}_i$. Transformation $\mathscr{T}_2$ computes suitable loudspeaker signals from the pre-filtered transformed signal components. This transformation is derived from the actual applied auralization algorithm. It should take care that the prescribed wave field $\tilde{\mathbf{w}}$ is auralized in the best possible way for a given loudspeaker setup. Wave field extrapolation, as described in the previous section, can be used for this purpose. Block $\mathscr{T}_3$ transforms the microphone array signals $\mathbf{l}$ into their transformed representations $\tilde{\mathbf{l}}$. Implementations of the plane wave decomposition for different microphone array geometries can be found in [4]. The error signal $\tilde{\mathbf{e}}$ describes the deviation of the auralized wave field from the reflection free case. The reflection free system response is modeled by the matrix $\tilde{\mathbf{F}}$ as for the traditional approaches. The error signal is then used to adapt the coefficients of the room compensation filters.

4. RESULTS

In order to verify our novel concept of active room compensation we performed simulations of a spatial audio system located in a typical listening room. We will first introduce the spatial audio system used and will then show some simulation results.

4.1 Wave Field Synthesis

Wave field synthesis (WFS) is a multichannel audio reproduction technique which allows a physically correct reproduction of wave fields. This section will only provide a brief introduction into WFS, a more detailed discussion can be found e.g. in [5].
WFS is based on the Huygens' principle. However, its mathematical foundation is given by the Kirchoff-Helmholtz integral. This basic principle can be used, similar as for wave field analysis, to synthesize a wave field within a volume V by setting the appropriate pressure distribution and its gradient on the surface. However, two essential simplifications are necessary to arrive at a realizable system: Degeneration of the surface S to a line and spatial discretization. Performing these steps the so called Rayleigh integrals can be derived [6]. The Rayleigh I integral states that a pressure field may be synthesized by means of a monopole distribution on a line. Using this result a WFS system can be realized by mounting closed loudspeakers in a linear fashion (linear loudspeaker arrays) surrounding the listening area leveled with the listeners ears. However, besides auralization, a WFS system can also perform other tasks. Because it provides control over the wave field inside the listening area, it can also be used to perform active room compensation in our context.
The fact that loudspeakers can only be mounted at discrete positions results in spatial aliasing. Above this aliasing frequency there is no control over the reproduced wave field. As a result, active room compensation can only be applied up to the aliasing frequency of the particular WFS system used. The aliasing frequency of a WFS system with a typical loudspeaker spacing of about $\Delta x = 19$ cm is $f_{al} \approx 900$ Hz [6]. Regarding the standard audio bandwidth of 20 kHz spatial aliasing seems to be a problem for practical WFS systems. Fortunately, the human auditory system seems not to be very sensitive to these aliasing artifacts so that the aliasing frequency plays no significant role for auralization. For active room compensation, passive damping methods in the listening room provide a solution to this problem. Passive methods for higher frequencies are much easier to realize than for low frequencies. Active listening room compensation could therefore be easily complemented by passive methods above the aliasing frequency.

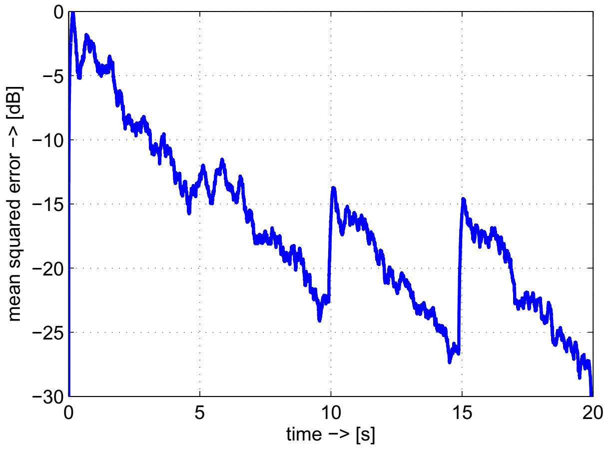

Figure 3: Resulting mean squared error between the compensated and the desired wave field.

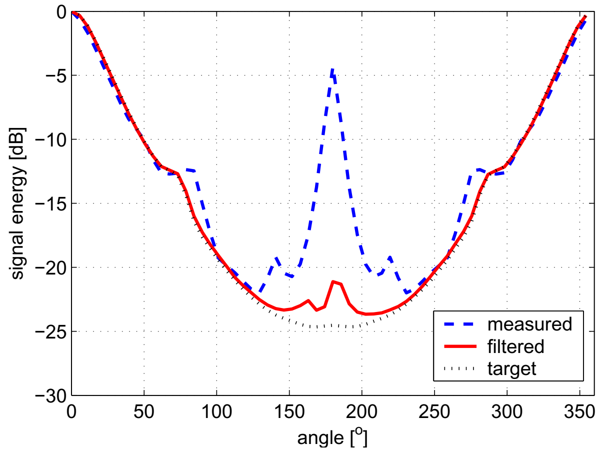

Figure 4: Room compensation results after convergence for the $\alpha = 0^{\circ}$ position shown as signal energy of the plane wave components.

4.2 Simulation results

The simulated setup equals the WFS setup at the demonstration room of our laboratory. The simulated active room compensation system consists of a circular loudspeaker array with diameter $D_{LS} = 3$ m with 48 equidistantly positioned loudspeakers. The loudspeaker array is placed in the center of the listening room with the size 6 m $\times$ 6 m $\times$ 3 m at a height of 1.80 m. The effect of the listening room on the auralized wave field was simulated using a 3D mirror image model. The reflection coefficient at the walls of the listening room was $R_{wall} = 0.5$, the reflection coefficient of ceiling and floor equaled zero. A circular microphone array with a diameter of $D_{Mic} = 2.90$ m is placed concentric inside the loudspeaker array. The microphone array consists of virtual pressure and pressure gradient microphones. The entire system was implemented as illustrated in Figure 2. Transformation $\mathscr{T}_1$ was chosen to create the representation of a virtual point source in the k_θ-domain. Transformation $\mathscr{T}_2$ extrapolates the pre-filtered signals to the loudspeaker positions utilizing equation (5). Transformation $\mathscr{T}_3$ calculates the plane wave decomposition according to [4] followed by an angular Fourier transformation (7). The compensation filters where calculated adaptively using the frequency domain adaptive filtering algorithm described in [7]. All signals were downsampled according to the aliasing frequency of the system of $f_{al} \approx 900$ Hz. To illustrate the performance of the proposed algorithm we performed a simulation with a moving virtual source. The virtual source is placed at a distance of 3 meters at $\alpha = [\ 30^{\circ}\quad 20^{\circ}\quad 10^{\circ}\quad 0^{\circ}\ ]$. The virtual source position is changed every 5 seconds.

Figure 3 shows the mean squared error of all plane wave components in $\tilde{\mathbf{e}}$ over the time axis corresponding to a omni-directional pickup of the error signal at the origin. It can be seen clearly that the algorithm is capable of adapting fast to the changed source positions. This is the benefit of the decoupling. Also calculated were the impulse responses of the target field, the measured and the compensated field in the plane wave domain after the adaption has converged for the last position. Figure 4 shows the signal energies of these impulse responses. The desired maximum energy at $\theta = 0^{\circ}$, resulting from the virtual source, as well as the reflections from the listening room can be seen clearly. The energy of the compensated wave field is nearly identical to the desired wave field. This illustrates the successful application of our proposed method.

5. CONCLUSION

The presented approach to the adaptive compensation of listening room reflections meets the requirements set out in Sec. 1.3. Based on proven theoretical methods from acoustic imaging and seismic wave theory, a generic technique has been developed which is suitable for application with emerging multichannel spatial audio systems. The presentation has been focused on the theoretical foundations and avoided associations with specific methods. Consequently, a number of further problems has not been addressed so far:

The description has been restricted to two spatial dimensions, which is an appropriate approximation for several practical cases. An extension to three spatial dimensions is straightforward but it requires a more elaborate notation (plane wave decomposition with respect to three spatial coordinates, plane wave extrapolation with respect to two spatial angles). Furthermore, the number of audio channels and the associated instrumentation expense increase considerably.

The presentation has been restricted to frequencies below the aliasing frequency, relying on the effectiveness of passive compensation methods. However, clever exploitation of aliasing effects may also facilitate active compensation above the aliasing frequency.

Finally, the deployed adaptive method (frequency domain adaptive filtering) has been addressed only by reference to [7]. Recent advances for adaptive inverse modeling methods will also support the application of the presented listening room compensation technique.

REFERENCES

[1] S. Spors, A. Kuntz, and R. Rabenstein, “An approach to listening room compensation with wave field synthesis,” in *AES 24th International Conference on Multichannel Audio*, Banff, Canada, June 2003, Audio Engineering Society (AES), pp. 49–52.

[2] J.A. Scales, *Theory of acoustic imaging*, Samizdat Press, 1997.

[3] A.J. Berkhout, *Applied Seismic Wave Theory*, Elsevier, 1987.

[4] E. Hulsebos, D. de Vries, and E. Bourdillat, “Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis,” in *110th AES Convention*, Amsterdam, Netherlands, May 2001, Audio Engineering Society (AES).

[5] S.Spors, H.Teutsch, A.Kuntz, and R.Rabenstein, “Sound field synthesis,” in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y.Huang and J.Benesty, Eds. Kluwer Academic Publishers, 2004.

[6] A.J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” *Journal of the Acoustic Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[7] H.Buchner, J.Benesty, and W.Kellermann, “Multichannel frequency-domain adaptive algorithms with application to acoustic echo cancellation,” in *Adaptive signal processing: Application to real-world problems*, J. Benesty and Y. Huang, Eds. Springer, 2003.