VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

Edice PhD Thesis, sv. 780

ISSN 1213-4198

Ing. Hasan Khaddour

Localization and Rendering of Sound Sources in Acoustic Fields

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

Ing. Hasan Khaddour

LOCALIZATION AND RENDERING OF SOUND SOURCES IN ACOUSTIC FIELDS

LOKALIZACE A INTERPRETACE ZDROJŮ ZVUKU
V AKUSTICKÝCH POLÍCH

Zkrácená verze Ph.D. Thesis

Obor: Teleinformatika
Školitel: Ing. Jiří Schimmel, Ph.D.
Oponenti: Ing. Martin Rakús, Ph.D.
Dr. Ing. Libor Husník
Datum obhajoby: 29. května 2015

Keywords
Multiple sound sources localization, DirAC, sound rendering, acoustic zooming.

Klíčová slova
Lokalizace zdrojů zvuku, DirAC, reprodukce zvuku, akustický zoom.

Místo uložení práce:
Disertační práce je k dispozici na Vědeckém oddělení děkanátu FEKT
VUT v Brně, Technická 10, Brno, 616 00

ISBN 978-80-214-5219-0
ISSN 1213-4198

Contents

Introduction

This dissertation combines two important parts of the acoustic discipline; namely, localization and rendering of sound sources. Both parts have been intensively investigated in the past decade. Although those two fields, localization and rendering, are studied separately, they are connected to each other. Where the remarkable ability of human beings to observe the surround environments can be seen as a common factor between these two fields i.e., the same cues, which the people use to localize the sound sources, are used in sound source localization methods and also are attempted to be re-created in the surround sound systems.

Surround sound systems have been the focus of attention for many years. New methods for spatial sound rendering are constantly appearing. They can be mainly used in cinemas, music, and video games. The spatial sound techniques not only provide the possibility of orientation, but also improve the pleasure of the listening to the sound. The aim of the modern spatial sound rendering methods is to bring the pleasure and the sensation of the musical places, such as the theaters and the opera houses, to the domestic sound systems.

The source localization methods are also used in many applications. They are primarily used in RADAR (Radio Detection and Ranging) and underwater SONAR (Sound Navigation and Ranging). Sound source localization in the air, however, is a quite new application of this technique compared to SONAR. Even so, it found its role in many applications in the modern technologies, for instance, in surveillance, speech recognition and teleconferencing.

This work deals with both localization and rendering methods, it aims at proposing a new method for sound source direction estimation, combining this method with an acoustic zooming technique and designing a new system that provides the possibility of sound source direction estimation and acoustic zooming in the same time. It should be noted that evaluation is an important part of each system. Therefore, throughout this work, some sound localization methods were evaluated and compared to each other, listening tests were also performed in order to compare the performance of existing sound spatial rendering techniques, and to choose the rendering technique with the best results to be used in the proposed acoustic zooming system.

Objectives of Dissertation

The main goal of the dissertation is to design and test an acoustic zooming system, which can locate and zoom the sound of one speaker among the multiple speakers.

In order to design such system, this work had to go gradually through several steps, which can be summarized as follows:

- Explore the existing surround sound techniques: The goal of this step is to choose the best sound rendering method, which can be used in the proposed acoustic zooming system.

- Investigate sound source direction estimation methods: This step introduces several sound source localization techniques, shows their advantages and disadvantages, and studies the absolute angle error of these techniques.

- Design a new sound source direction estimation method and evaluate the proposed method: The goal of this step is to suggest a new method, which can be used in the proposed acoustic zooming system in order to estimate the direction of multiple speakers.

- Implement an acoustic zooming system using DirAC and sound source direction estimation method: The purpose of this step is to propose a system which can estimate the direction of arrival of multiple speakers and also choose the sound of one of them and zoom it while attenuating the other sounds.

- Perform the subjective and objective quality assessment to evaluate the proposed acoustic zooming system: Evaluation is an essential part of each system. Therefore, this step aims at evaluating the proposed system by performing the listening tests and applying objective measurements using several time-frequency transforms.

Chapter 1

A Comparison and Evaluation of Localization and Rendering Methods

This chapter investigates the accuracy of some sound source localization methods, and also provides the listening tests to evaluate and compare the performance of sound reproduction techniques.

In order to compare the performance of sound source localization methods, the methods were first simulated in Matlab, and then the measurements were performed in the laboratory to affirm the simulation results.

The evaluation of sound rendering methods is proceeded by performing the listening tests in laboratory. The listening tests were designed to compare the average absolute angle error of these methods when the listener is moved out of the center of the loudspeaker array.

1.1 Comparison of Sound Source Localization Methods

In order to evaluate the time delay of arrival based methods, we simulated in Matlab three methods; namely, cross-correlation (CC), phase transform (PHAT) and maximum likelihood (ML), and we studied the impact of the existence of the noise signal on their performance [1]. The experimental results are also presented in this section.

1.1.1 Simulation Results

The simulation process was carried-out in Matlab. A normal male voice was chosen as a sound source. Gaussian noise was used as an additive noise. Two signals were used and one of them was delayed and then both signals were precessed using the mentioned methods.

Figure 1.1 shows the simulation results when the noise signals are added to the original signal. As can be clearly seen, the three methods were able to estimate the time delay of arrival correctly. Another peaks appear in the simulation. However, the highest peak still indicates the right time delay estimation. It can be also seen that the PHAT method achieved a sharper peak than the other methods. The maximum likelihood method also achieved a sharp peak, but it has several additional smaller peaks. Although cross-correlation has the widest peak, it still can estimate the real time delay in the simulation conditions. It should be also noted that PHAT method has the biggest peak denoting the time delay estimation compared to the additive peaks around it, whereas the other methods have bigger additional peaks during the time interval caused by the additive noise [2].

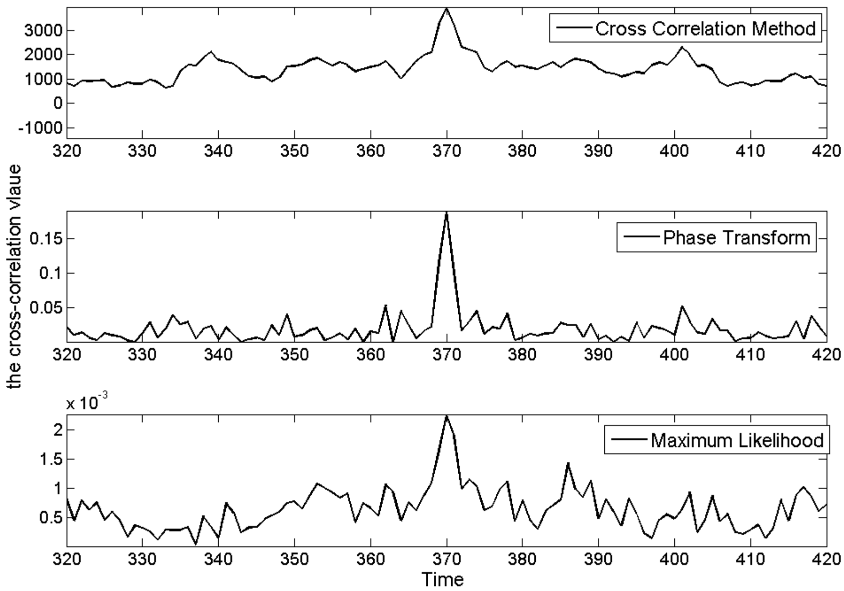


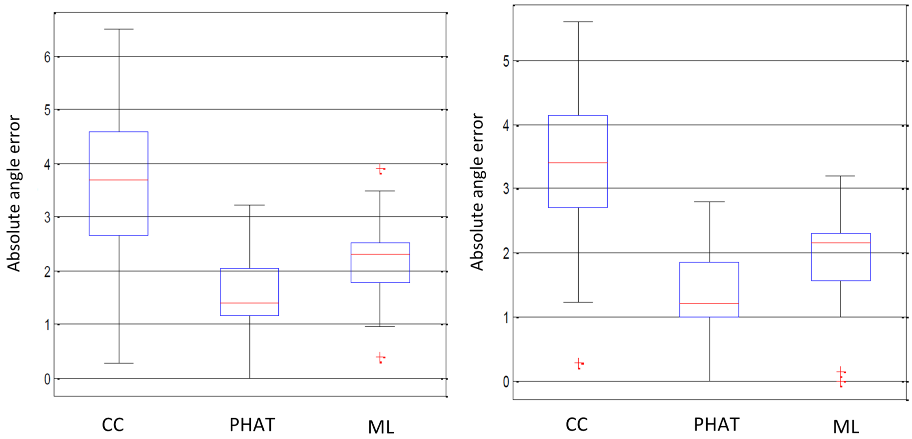
Figure 1.1: The simulation results of CC, PHAT, and ML methods with an additive Gaussian noise.

1.1.2 Experimental Results

The experiment was performed in an acoustic laboratory as follows: In the first part of our experiment, two microphones were located in the middle of the laboratory; the distance between the microphones was chosen to be 16.5 cm. In the second part, we used three microphones. In the both parts, the sound was recorded in thirty six different angles in the front side of the microphone array. These angles were equidistantly separated (i.e., 5 degrees from each other).

However, the experimental results were expected regarding the simulation results that were shown in section 1.1.1. The error in the results is caused because of the noise, reverberant signals, and the correlation between these signals and the original sound signal.

As can be seen in Figure 1.2b, adding the third microphone did not actually improved the direction of arrival estimation when the sound sources were located in the front side of



(a) Two microphones. (b) Three microphones.

Figure 1.2: The average absolute angle error when two and three microphones are used.

the microphone array. However, it provides extra information that can be used for sound source localization, i.e., the distance of the microphone array.

This information about the sound source position can be derived using so-called *triangulation* once the time delay of arrival is estimated [3].

1.2 Localization Blur of 2D Ambisonic and VBAP

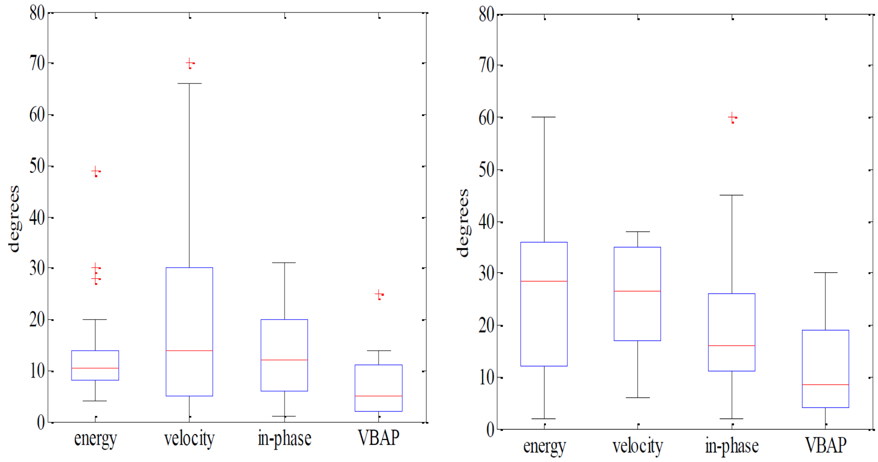
A part of this work is devoted to design a rendering system with acoustic zooming. Therefore, an investigation about the most suitable rendering method has been made. Three Ambisonic decoders have been tested; namely, energy decoder (max r_E), velocity decoder and in-phase decoder.

1.2.1 Experimental Results

The results are illustrated using box plots. The boxes show the average absolute angle error between the position we wanted the virtual sound to be, and the position that the listener perceives the sound to be coming from. The results are shown for each decoder in the three positions the listen was seated in.

At sweet spot, the Ambisonic decoders have similar median error, see Figure 1.3a. The interquartile range is the best for the energy decoder and the worst for the velocity decoder with the in-phase decoder in the middle. But the best localization is achieved using the vector base amplitude panning method.

The performance of the positioning methods decreased at the position 0.25 m away from the center, see Figure 1.3b. Among Ambisonic decoders, the in-phase decoder was the least affected by the movement. The VBAP method has the best localization accuracy at this position.



(a) Central position. (b) 0.25 m far from the center.

Figure 1.3: The average absolute angle error of the three Ambisonic decoders and VBAP.

At the position 0.5 m away from the sweet spot the median error of the Ambisonic decoders is almost identical to each other, see Figure 1.4. The velocity decoder has a significant interquartile range meaning the measured values had the most spread at this position. The in-phase decoder's interquartile range remains unchanged. The VBAP method proved to be the best positioning method for this position.

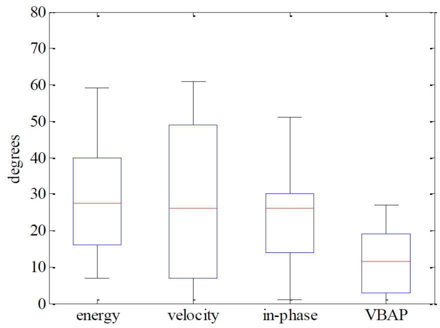


Figure 1.4: The average absolute angle error of the three Ambisonic decoders and VBAP at 0.5 m far from the center.

The results show that VBAP has a better localization performance at all positions than first-order Ambisonic. Both methods achieved their best at the sweet spot. However, the listeners tend to judge the direction of the VBAP virtual sound source to be more in the direction of the loudspeakers as they are sitting far from the sweet spot, whereas the Ambisonic virtual sound source become wider and sometimes the listeners can recognize two or more correlated sounds coming from different distances. VBAP produces virtual sound source that is sharper than virtual sound source produced by Ambisonic [4].

Chapter 2

Estimation of the Direction of Arrival of Multiple Speakers

This chapter introduces a new method for sound source direction estimation. The new method is called Energetic Analysis Method. It can be used for sound source direction estimation in both two dimensional and three dimensional plane. It can be also used for tracking a speaker in the horizontal plane. Simulation and the experimental results of this method approved its accuracy. Thus, we will used it later as a compatible method with acoustic zooming system.

2.1 Energetic Analysis Method

Energetic analysis method is a technique for sound source direction estimation, based on analyzing B-format signals [5]. This method is able to estimate the direction of multiple speakers in both two and three dimensional plane.

2.1.1 Simulation Results in Horizontal Plane

The method is simulated in Matlab using different time-frequency transforms, namely Gabor transform, STFT and zero padding. Figure 2.1 illustrates the simulation results when a real noise is used. As can be seen, the method was able to estimate the direction of arrival of the simulated sound sources. However, when STFT and zero-padding were used, additive peaks are noticed and that can lead to estimate the direction of fake sound sources that belong originally to noise sources. However, this effect was less noticeable when Gabor transform was used. That can be explained by the fact that Gabor transform achieves better resolution in time-frequency representation.

The simulation results showed the ability of this method to estimate the direction of arrival of multiple sound sources in the horizontal plane. It was shown that better resolution in time-frequency representation ensures better accuracy for this method [6].

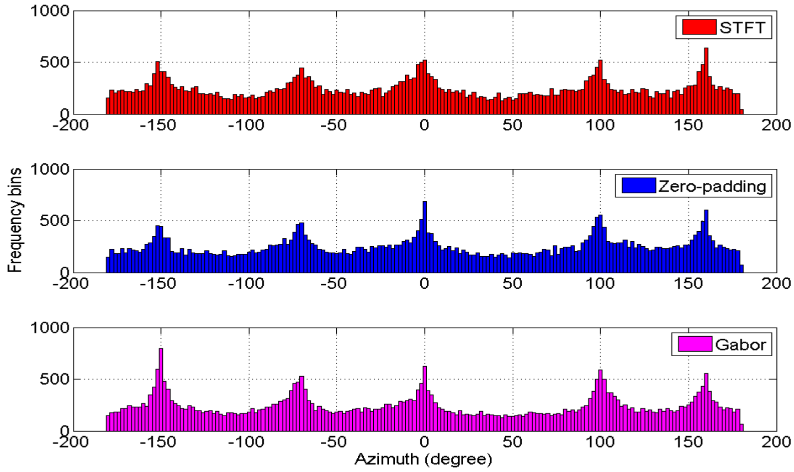


Figure 2.1: Simulation result with the presence of an additive real noise.

2.1.2 Simulation Results in Three Dimensional Plane

In order to estimate the direction of arrival of multiple sound sources in three dimensional space, the four B-format signals are needed.

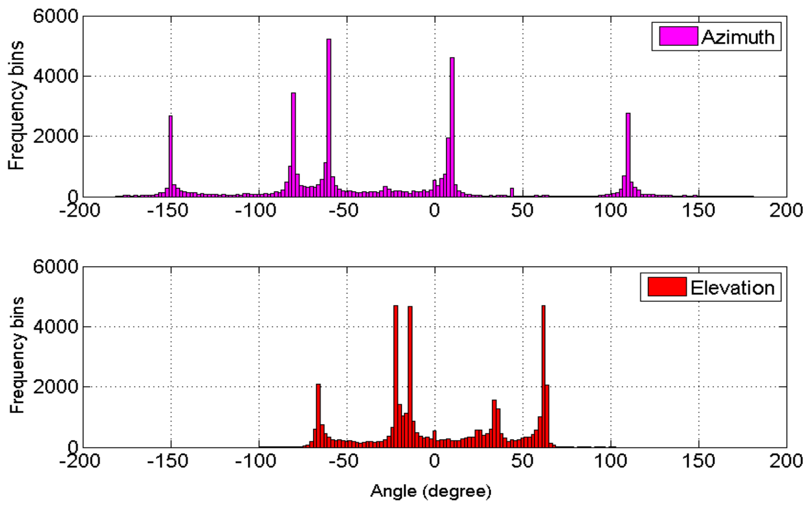


Figure 2.2: Simulation result in three dimensional plane with an additive noise.

Figure 2.2 shows the simulation results for this method in three dimensional plane. As can be seen in Figure 2.2, the peaks denote the direction of the sound sources in both horizontal and vertical plane, and the method was able to estimate the direction of multiple speakers in 3D [7].

2.1.3 Experimental Evaluation of Energetic Analysis Method

In order to evaluate the sound direction estimation method, experiments were performed in real environments. The experiments have been carried out for both two-dimensional and three-dimensional plane.

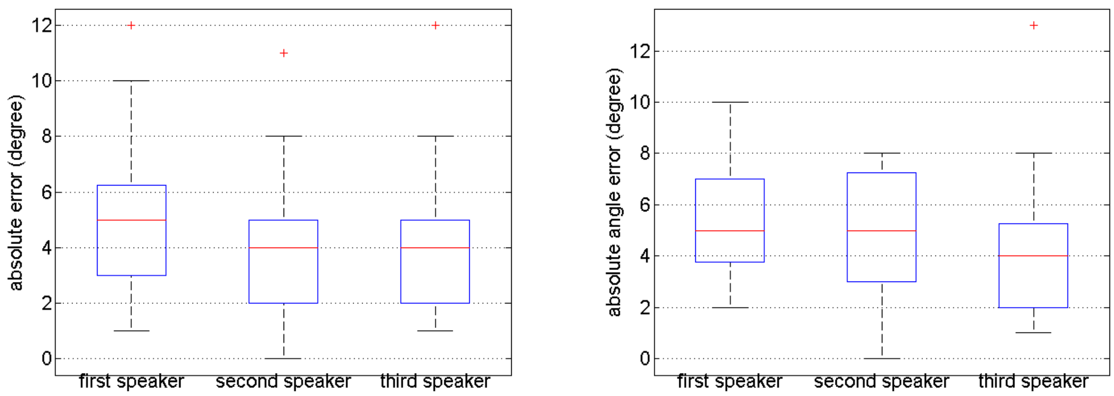
Experimental Results in the Horizontal Plane

The method was applied to the recorded sound files using different time-frequency transform, which ensures the possibility of comparison of the results depending on the resolution of the time-frequency transform in the time-frequency plane.

Figure 2.3a shows the result in the horizontal plane when STFT was applied. As can be seen from this figure, the method was able to estimate the direction of the speakers in the horizontal plane. The median absolute angle error in this case was about 4 degrees.

Experimental Results in the Vertical Plane

The experimental results in the vertical plane are showed in Figure 2.3b. The median absolute error was almost 4 degrees. The method was able to estimate the direction of the speakers in the vertical plane. However, the median error in this case was bigger than the median error in the horizontal plane.



(a) Horizontal plane.

(b) Vertical plane.

Figure 2.3: The average absolute angle error.

Chapter 3

Acoustic Zooming

In this chapter, we present a compatible system for both acoustic zooming and sound direction estimation [8]. The proposed system relies on the energetic analysis method for estimation the direction of arrival of multiple speakers, and on the changing the parameters of the directional audio coding for zooming the sound coming from wanted direction.

The listening tests have been implemented to evaluate the reliability of this system, the tests were designed to compare the quality of the sound when several time-frequency transforms were used, and to evaluate the ability of this system to zoom the sound of one speaker and attenuate the other. The experimental results showed the ability of this system to zoom the wanted sound. Objective measurements have been carried out as well. The results of the measurements showed the quality of the proposed system.

3.1 Description of the Proposed System

The proposed system uses the energetic analysis method to estimate the direction of multiple speakers giving the information about the speaker locations, which is needed in the zooming system [8]. Then, the system zoom the sound of one speaker by modifying the parameters of the directional audio coding. Further more, after estimation the direction of the speakers, the system can be modified to attenuate the sound of one or more speakers and keep the sound of the others at the same level as in the original sound.

The proposed system consists of four blocks; namely, sound source localization unit, DirAC analysis unit, zooming and synthesis unit and rendering unit, see Figure 3.1.

The input signals for this system are B-format signals, the output signals are the modified DirAC signals.

3.1.1 The Modified Energetic Analysis Method

The energetic analysis method has been introduced in chapter 2. However, extra steps have been added to the original energetic analysis method to exploit the proprieties of the

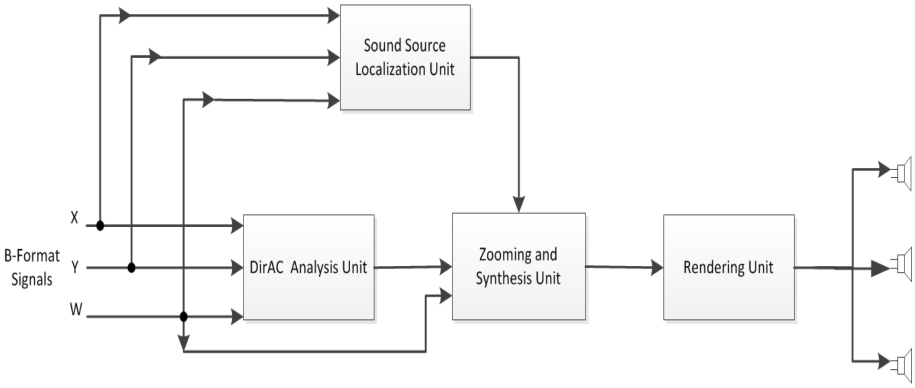


Figure 3.1: The proposed system in the two-dimensional plane.

human voice and also the proprieties of the acoustic in the closed room, see Figure 3.2. These three steps are: filters, DirAC analysis and estimation of the non-diffuse part.

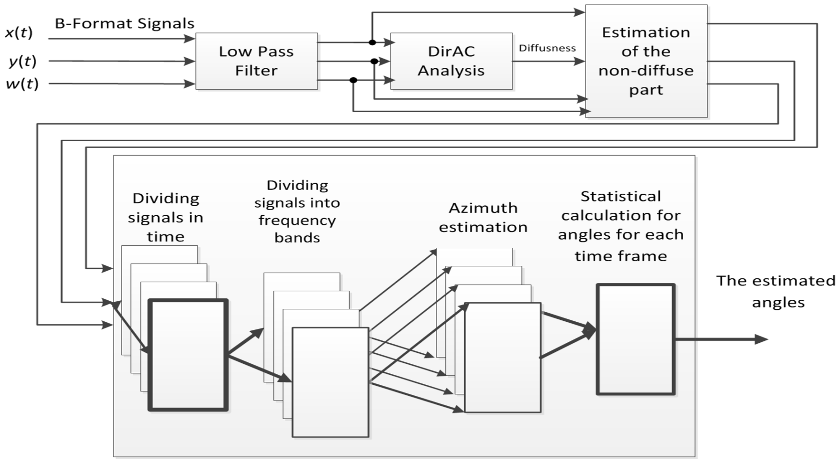


Figure 3.2: The localization unit in the two-dimensional plane.

Filters

The input signals of this unit are B-format signals. The idea of using a low pass filter comes from the fact that we want to estimate the direction of a human speech source. As it is well known, the speech spectrum can be divided into two parts, the first part is flat and it contains the frequencies up to 500 Hz, whereas the second part has a slope of −10 dB/octave, and it is applied to the frequencies higher than 500 Hz [9], [10]. Applying a low pass filter to the input signals suppresses the additional interference caused by higher frequency, which belongs to the noise signals. Therefore, we applied a low pass FIR filter with cut-off frequency equals to 3500 Hz.

We also applied a high pass filter with cut-off frequency equal to 100 Hz in order to minimize the effect of the standing waves in the laboratory. It was seen that adding these filters improves the accuracy of the energetic analysis method.

Dirac Analysis

The purpose of DirAC analysis in this stage is to estimate the diffuseness parameter, which can be used to divide the sound signal into diffuse and non-diffuse streams. The input signal of this step is the resulted filtered signals from the previous step. The signals are then divided in time and frequency, and the DirAC parameters are calculated as in the original DirAC method. The diffuseness parameter is then estimated to be applied in the next step.

Estimation of the Non-Diffuse Part

The sound signals are first separated into diffuse and non-diffuse stream using the diffuseness parameter. The separation can be done by multiplying the signal in the frequency domain by the parameter $\sqrt{\Psi}$ and $\sqrt{1-\Psi}$. Then, the non-diffuse part can be used to improve the accuracy of this unit by eliminating the diffusing sound, which is resulted from the reverberant sound. The non-diffuse part only is transmitted to the time domain using inverse STFT or Gabor transform depending on the transform used in the transformation into frequency domain.

After processing the above mentioned steps, the original energetic analysis method is applied normally to the resulted signals. The results are in this case more accurate because of suppressing the interference caused by the diffuse sound and reverberant signals.

3.1.2 Zooming and Synthesis Unit

The input signals for this unit are the omni-directional B-format signal ($w(t)$), the parameters estimated from the DirAC analysis unit and the information about the direction of the speakers obtained from the sound source localization unit [8].

The sound signal is first transmitted into frequency domain, and then it is divided into diffuse and non-diffuse stream depending on the diffuseness we estimated from the DirAC analysis unit. A gain factor is then applied to the non-diffuse part, and it is calculated as

$$g(m,n) = \begin{cases} g_{max} & \text{if } \text{DOA}(m,n) \in [\theta+\Upsilon, \theta-\Upsilon] \\ \\ g_{min} & \text{if } \text{DOA}(m,n) \notin [\theta+\Upsilon, \theta-\Upsilon] \end{cases} \tag{3.1}$$

where $g(m,n)$ is the gain applied to the frequency bin number m in the time sample number n, g_{max} is the maximum gain applied to the sound we want to zoom, g_{min} is the attenuation factor, $\text{DOA}(m,n)$ is the direction of arrival estimated from DirAC analysis, θ is the direction of the speaker whose sound we want to emphasis, and it is estimated from the sound source localization unit and Υ is the half of the angle in which we zoom the sound and it differs in each scenario. Υ was chosen to be 5 degrees in our experiments.

It was chosen depending on the length of the arc (space) that the normal size person can occupy when he is 2 m far from the microphones.

The zooming factor impacts the quality of the sound. When a large zooming factor is used, an audible distortion occurs to the sound file, which affects the quality of the reproduced sound. Using a smoothing method improves the quality of the sound, and minimizes the distortion of the sound.

3.1.3 Rendering Unit

When the sound is transmitted to the time domain, it can be rendered to a set of loudspeakers, or to headphones [11]. However, a prior knowledge about the distribution of the loudspeakers should be taken into account when the rendering method is applied. In our system, we chose VBAP as a suitable method for rendering the sound since it has better localization accuracy over first-order Ambisonic [12].

The experiments were designed to evaluate the ability of zooming the sound, the resolution of the zooming technique and the precision of the mentioned system. They can be divided into three stages; namely, recording the sound, processing the sound and listening stage [8].

3.2 Listening Test

In order to evaluate the zooming system, a listening test was carried out [8]. The listening test compared the original sound rendered using DirAC and the zoomed sound using both STFT and Gabor transform. The test was performed in the acoustic laboratory as follows: six loudspeakers were located in the vertices of a regular hexagon with distance of vertices from the sweet spot of 2.5 meters. For this test, ten listeners were used. The listeners have been chosen without any hearing impairment, at the age from 25 to 35 years. Five listeners have a good experience in the procedure of listening tests. For others, the procedures were explained carefully. The listeners included four women and six men. Each listener was seated at the position of the sweet spot of the loudspeaker setup. The listeners were asked to give an evaluation of the quality of the sound and of the loudness of the loudest speaker compared to the others. They were told to write their evaluation on a sheet of paper, which had the questions and a scale for each question. Five scales were available to describe the quality of the sound based on mean opinion score (MOS) [13]. The existed options according to MOS are presented in Table 3.1.

Another five scales were available to describe the loudness ratio of the speakers to each other. The available options that describe the ratio of the loudness of the speakers in this case are shown in Table 3.2.

Quality of the speech	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Table 3.1: Listening-quality scale (MOS).

The loudness ratio of the loudest speaker	Score
I cannot hear the others	5
Very higher	4
Higher	3
Slightly higher	2
The speakers have the same loudness	1

Table 3.2: Loudness ratio.

The listening tests were also devoted to measure the precision of the system. The listeners were asked to localize the sound sources. A mobile loudspeaker was used as a reference sound. The same sentence was rendered via the mobile loudspeaker and the original loudspeaker array alternately. The mobile loudspeaker was moved around the sweet spot in the same distance as the loudspeakers of the array till the listener said that the sound coming from it and the sound rendered via the original loudspeaker array have the same direction. This step was applied to each one of the four speakers in each audio file and only to the zoomed speaker in the zoomed files.

It should be noted that the duration of each test did not exceed 30 minutes, during which each listener evaluated three sound files.

3.3 Experimental Results

A part of our experiments attended to measure the localization blur of this system, and the influence of the zooming system on this blur. In our experiments, most of the listeners explained the sound localization as (*easier*) when the zooming was applied. However, it was noticed that the listeners attended to match the sound source with the visible loudspeakers when the sound source was near them. In the original sound files i.e. without zooming, the listeners were asked to localize the four speakers, whereas they were asked to localize only the zoomed sound when the zooming was applied. The results showed that the median blur for the system was about 18°, and it was decreased a little bit when the zooming sound was applied. This little improvement in precision is mostly because of attenuating of the other sounds, which can be seen as a distraction when the listener focuses his attention on one speaker, see Figure 3.3.

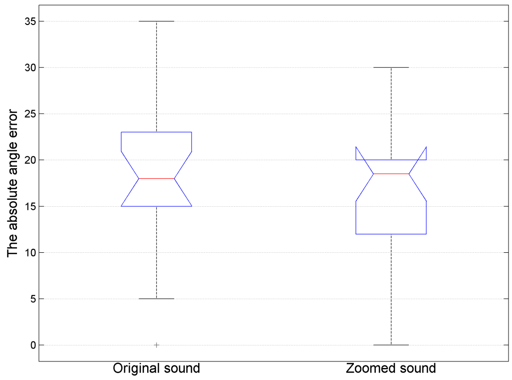
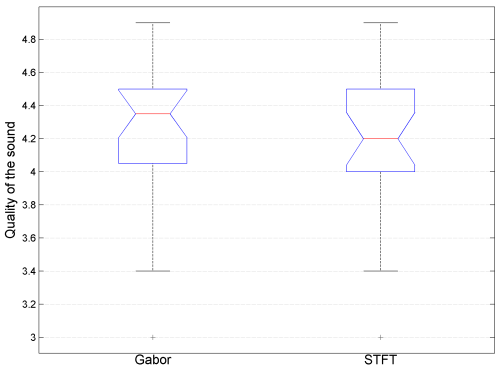
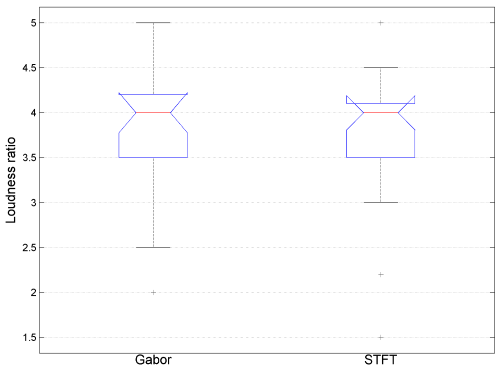


Figure 3.3: The localization blur for both original and zoomed sound.

Depending on our results, the experienced listeners, who had taken a part in listening tests before, felt the difference between the quality of the sound files when Gabor or STFT was applied to the zooming system more than the inexperienced listeners. Figure 3.4a and Figure 3.4b compare the results when Gabor and STFT are used using boxplot. As can be clearly seen, Gabor achieved the best results. The perception of the loudness of the sound was better and it kept better sound quality.



(a) The quality of the sound.



(b) The loudness ratio.

Figure 3.4: Experimental results when Gabor and STFT are used.

3.4 Objective Measurement

For the objective assessment of quality of extracting the signal of the zoomed sound source from a given mixture we used the PEASS algorithm [14] which is designed specifically for these purposes. The algorithm [14] is based on decomposing the estimation error into three components (target distortion, interference and artifacts components), assessing the salience of each component via PEMO-Q

We used the sound of four male speakers as the source sounds of the zooming system. The sound of the speakers was recorded in the anechoic room (reverberation time 50 ± 10 ms in octave bands from 250 Hz to 8 kHz). The sampling frequency of the recordings was 44.1 kHz and the recordings were synchronized in time. In order to align the loudness of the sound sources, their level was adjusted to RMS value of -20 dBFS with maximum peak values of -3 dBFS using the Steinberg Wavelab loudness normalizer. These recordings were rendered using four loudspeakers in the same laboratory where the subjective tests were performed. The loudspeakers were placed in the same distance from the sweet spot and in the same angles as the speakers when the recordings for the subjective tests were carried out.

The results of the objective assessment of the speech quality are shown in Table 3.3. As it can be seen from the PEASS results, the overall perceptual score of the zoomed speaker is definitely better than the score of all four speakers played back simultaneously (OPS=8) and also better than the score of the sound field of all four speakers rendered using DirAC without zooming (OPS=19). The results are almost the same when Gabor and STFT are used for the zooming. A more detailed analysis shows that a greater suppression of the other speakers (IPS) occurs using the Gabor transformation than the STFT.

Tested signal	Original sound field of 4 speakers	Sound field rendered using DirAC without zoom	Sound field rendered with zoom	
			Gabor	STFT
OPS	8	19	38	38
TPS	81	38	44	44
IPS	1	15	55	52
APS	87	54	44	45

Table 3.3: The average results of the speech quality assessment using the PEASS algorithm.

Chapter 4

Conclusions

This dissertation aimed at designing an acoustic zooming system, which enables zooming the sound of one speaker among the other speakers. Throughout this work, two major disciplines have been investigated; namely, sound source localization and surround sound rendering. The proposed system uses the results of this work in the both mentioned fields, creating the possibility of further investigation in this area in order to work in the real time.

Contributions of the Thesis

The main contributions of this work can be summarized as follows:

- The thesis started by investigating the existed sound source localization methods. Some experiments have been implemented in order to compare the accuracy of these methods.
- A new sound source direction estimation method, which was inspired by DirAC, has been introduced in this work. The simulation and the experimental results of this method have been shown in Chapter 2.
- A system for acoustic zooming has been tested using several time-frequency transforms. This system depends on the sound direction estimation method and on modifying the parameters of DirAC.
- The listening tests and objective measurements have been performed as well. The results of these listening tests and measurements approved the accuracy of this acoustic zooming system.

The author of the thesis suggests exploring the possibility of working the proposed system in the real time. Then, the system can be used in video-conferencing system.

Bibliography

[1] H. Khaddour, "A comparison of algorithms of sound source localization based on time delay estimation," *Elektrorevue,Internetovy casopis*, vol. 2, no. 1, pp. 31–37, 2011.

[2] H. Khaddour and A. Warda, "Sound source localization using time delay estimation." in *6th International Conference on Teleinformatics ICT.*, 2011, pp. 179–182.

[3] Ø. Hjelle and M. Dæhlen, *Triangulations and Applications*, ser. Mathematics and Visualization. Physica-Verlag, 2006.

[4] M. Trzos and H. Khaddour, "Localization blur of 2d ambisonic reproduction in small rooms." in *Telecommunications and Signal Processing (TSP), 2010 33th International Conference on*, August 2010, pp. 56–59.

[5] H. Khaddour and J. Schimmel, "Multiple sound sources localization using energetic analysis method," *Elektrorevue,Internetovy casopis*, vol. 3, no. 4, pp. 5–9, 2012.

[6] H. Khaddour and D. Kurc, "Impact of applied transform on accuracy of energetic analysis method," in *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*, July 2013, pp. 464–468.

[7] H. Khaddour, J. Schimmel, and M. Trzos, "Estimation of direction of arrival of multiple sound sources in 3d space using b-format," *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 2, no. 2, pp. 63–67, 2013.

[8] H. Khaddour, J. Schimmel, and F. Rund, "A novel combined system of direction estimation and sound zooming of multiple speakers," *Radioengineering*, vol. 24, no. 2, 2015, in print.

[9] N. Wittenberg, *Understanding Voice Over IP Technology.* Cengage Learning, 2009.

[10] S. Furui, *Digital speech processing: synthesis, and recognition.* CRC Press, 2000.

[11] V. Pulkki, "Applications of directional audio coding in audio," in *Proceedings of the 19th International Congress of Acoustics*, 2007.

[12] H. Khaddour and M. Trzos, "Representation of sound field using ambisonic," *Elektrorevue,Internetovy casopis*, vol. 1, no. 2, pp. 1–7, 2010.

[13] Methods for subjective determination of transmission quality. Accessed: 2013-12-5. [Online]. Available: http://www.itu.int/rec/T-REC-P.800-199608-I/en

[14] E. Vincent, "Improved perceptual metrics for the evaluation of audio source separation," in *Latent Variable Analysis and Signal Separation.* Springer, 2012, pp. 430–437.

Author's Publications

[1] H. Khaddour and M. Trzos, "Representation of sound field using ambisonic," *Elektrorevue,Internetovy casopis*, vol. 1, no. 2, pp. 1–7, 2010.

[2] M. Trzos and H. Khaddour, "Localization blur of 2d ambisonic reproduction in small rooms." in *Telecommunications and Signal Processing (TSP), 2010 33th International Conference on*, August 2010, pp. 56–59.

[3] H. Khaddour, "A comparison of algorithms of sound source localization based on time delay estimation," *Elektrorevue,Internetovy casopis*, vol. 2, no. 1, pp. 31–37, 2011.

[4] H. Khaddour and A. Warda, "Sound source localization using time delay estimation." in *6th International Conference on Teleinformatics ICT.*, 2011, pp. 179–182.

[5] H. Khaddour, "Sound source localization based on b-format signals," in *Telecommunications and Signal Processing (TSP), 2011 34th International Conference on.* IEEE, 2011, pp. 335–338.

[6] M. Trzos, and H. Khaddour "Efficient Spectral Estimation of Non- Stationary Harmonic Signals Using Harmonic Transform, " *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 1, no. 2- 3, pp. 1–3, 2012.

[7] H. Khaddour and J. Schimmel, "Multiple sound sources localization using energetic analysis method," *Elektrorevue,Internetovy casopis*, vol. 3, no. 4, pp. 5–9, 2012.

[8] H. Khaddour and D. Kurc, "Impact of applied transform on accuracy of energetic analysis method," in *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*, July 2013, pp. 464–468.

[9] H. Khaddour, J. Schimmel, and M. Trzos, "Estimation of direction of arrival of multiple sound sources in 3d space using b-format," *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 2, no. 2, pp. 63–67, 2013.

[10] D. Kurc, V. Mach, K. Orlovsky, and H. Khaddour, "Sound Source Localization with DAS Beamforming Method Using Small Number of Microphones," in *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*, July 2013, pp. 526–532.

[11] H. Khaddour, J. Schimmel, and F. Rund, "A novel combined system of direction estimation and sound zooming of multiple speakers," *Radioengineering*, vol. 24, no. 2, 2015, in print.

Curriculum Vitae

Personal Information

Name: Hasan Khaddour.
Title: Ing.
Place of Birth: Latakia, Syria.
Nationality: Syrian.
Telephone: +420 776 670 485.
Email : xkhadd00@stud.feec.vutbr.cz

Education

2009-2015: Ph.D. student at Brno University of Technology, Czech Republic.
2002-2007: Engineering student at Tishreen University, Syria.
1999-2002: Upper Secondary Education, Latakia, Syria.

Employment History

2007 - 2008: Assistant at Faculty of Mechanical and Electrical Engineering, Department of Communication and Electronic Engineering.

Languages

Arabic - native speaker.
English - excellent command.
Czech - very good command.

Honors and Awards

Top Students Award for five years in a row at Tishreen University, Syria.

Participation in Projects

- FEKT-S-11-17 – Research of Sophisticated Methods for Digital Audio and Image Signal Processing. Holder: Prof. Z. Smékal. 2011.
- MSM21630513 – Electronic Communication Systems and Technologies of Novel Generations (ELKOM). Holders: Prof. Z. Raida, Prof. K. Vrba, Prof. J. Jan. 2008–2011.
- 1595/F1/2012 Innovation of Studio Engineering subject. Holder: Ing. J. Schimmel, 2012.

Abstract

This doctoral thesis deals with sound source localization and acoustic zooming. The primary goal of this dissertation is to design an acoustic zooming system, which can zoom the sound of one speaker among multiple speakers even when they speak simultaneously. The system is compatible with surround sound techniques.

In particular, the main contributions of the doctoral thesis are as follows:

1. Design of a method for multiple sound directions estimations.

2. Proposing a method for acoustic zooming using DirAC.

3. Design a combined system using the previous mentioned steps, which can be used in teleconferencing.

Keywords:
Multiple sound sources localization, DirAC, sound rendering, acoustic zooming.

Abstrakt
Disertační práce se zabývá lokalizací zdrojů zvuku a akustickým zoomem. Hlavním cílem této práce je navrhnout systém s akustickým zoomem, který přiblíží zvuk jednoho mluvčího mezi skupinou mluvčích, a to i když mluví současně. Tento systém je kompatibilní s technikou prostorového zvuku.

Hlavní přínosy disertační práce jsou následující:

1. Návrh metody pro odhad více směrů přicházejícího zvuku.

2. Návrh metody pro akustické zoomování pomocí DirAC.

3. Návrh kombinovaného systému pomocí předchozích kroků, který může být použit v telekonferencích.

Klíčová slova:
Lokalizace zdrojů zvuku, DirAC, reprodukce zvuku, akustický zoom.