G_007

Arbitrary Listening-point Generation Using Acoustic Transfer Function Interpolation in A Large Microphone Array

Mehrdad Panahpour Tehrani † Yasushi Hirano† Shoji Kajita† Toshiaki Fujii †† Kazuya Takeda††† Kenji Mase †

1. Introduction

So far, several approaches have tried to generate arbitrary listening-point of sound; however there are few effective models such as Head Related Transfer Function (HRTF) [1] and representation of the sound sources in 3D space [2] to allow for an efficient processing. On the other hand, images are rendered by computer graphics algorithms and have become more attractive and more efficient and image synthesis hardware has come to existence, such as Free viewpoint TV (FTV) [3]. Therefore the corresponding sound in the virtual location of camera is necessary to be generated for future 3D audiovisual systems. In this research, the problem of arbitrary listening-point generation using the measured acoustic transfer function (ATF) is considered.

2. ATF Estimation

The ATFs from all dead sound sources (i.e. closely listened sound without crosstalk with other sources) to all far field microphones are estimated using a Time-Stretched Pulse (TSP) as shown in Fig 1 and Eq. 1.

$$ATF(\omega)=\frac{M(\omega)}{S(\omega)} \quad (1)$$

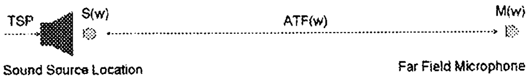

Fig.1: Sound source "*S(w)*" and far field sound "*M(w)*" recording of TSP signals for ATF estimation.

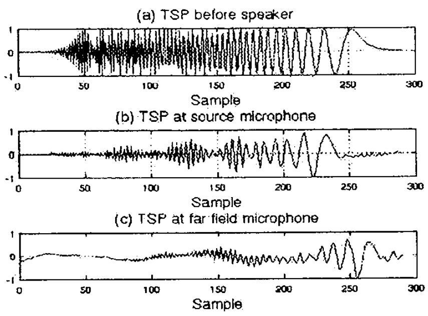

Fig.2: TSP signal, 256 sample, 48000KHz, 16 bit.

TSP for a short duration of time (e.g. 256 samples, 48000KHz, 16 bits) is broadcasted at the location of sound source to avoid the reflection of the environment.

Fig. 2 shows the TSP signal, recorded TSP signals at source, and far field microphone signal at nearly 4m distance to the source, respectively.

† Information Technology Center, Nagoya University
†† Graduate School of Engineering, Nagoya University
††† Graduate School of Information Science, Nagoya University

After generating all ATFs for all microphones, the arbitrary listening-point generation is performed after generating a pool of ATFs by the interpolation of ATFs of each pair of microphones as explained in the following.

3. Arbitrary Listening-point Generation using ATF-pool

For arbitrary listening-point generation, ATFs from all sound sources to the desired listening-point must be synthesized. Therefore, the sound in the desired point can be generated by convolving the close microphone sound data with the ATF for each sound source and then combining them.

3.1 ATF-pool

The ATF-pool contains all at microphones' locations, plus synthesized ATFs with a chosen density from sound sources to any point on lines/planes which pass all close pairs/Triple of microphones, as it is shown in Fig. 3.

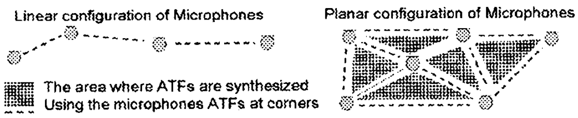

Fig.3: ATF-pool data configuration in microphone array.

The ATFs between pairs or on the plane among three microphones are synthesized by applying a weighted linear interpolation of the microphones' ATFs in frequency domain, as shown in Eq. 2 for both linear and planar configuration of microphone arrays.

$$ATF_k(\omega)=\sum_{i=1}^{n}\lambda_i ATF_{M_i}(\omega), \sum\lambda_i=1 \begin{cases} Linear \rightarrow n=2 \\ Planar \rightarrow n=3 \end{cases} \quad (2)$$

where symbol k is the ATF to be synthesized on line or plane. Symbol i refers to each microphone in the corner of the line or plane, and symbol λ_i is the weighting factor for linear interpolation of the coefficients of ATF's of microphones (M_i) at each corner.

Note that the denser the ATF-pool, the smoother an arbitrary listening-point can be generated when the arbitrary listening-point is roaming in the listening area.

3.2 Arbitrary Listening-point Generation

Given the ATF-pool, an arbitrary listening-point in a far field area (i.e. almost more than 1m distance from sound sources) can be generated as following for two cases as shown in Fig. 4.

Case 1: If the given listening-point is on the area of ATF-pool, the closest ATF (e.g. ATF(S1), ATF(S2)) to the point is chosen for each sound source. Then, the listening-point sound can be generated by convolving each source sound data with its ATF at the listening-point and then adding them.

Case 2: If the listening-point is located out the area of the ATF-pool, closest ATFs in the ATF-pool, which are on the line pass through the listening-point and source location, are chosen (e.g. ATF(S1), ATF(S2)). Then, the chosen ATFs in ATF-pool are compensated considering the attenuation in far field to the listening-point (e.g. ATF'(S1), ATF'(S2)). Finally, the sound of the given listening-point is generated in a similar way by convolving each source sound data with its ATF at the listening-point and adding them. Fig. 4 shows how the ATFs are chosen for an arbitrary listening-point for a linear microphone array.

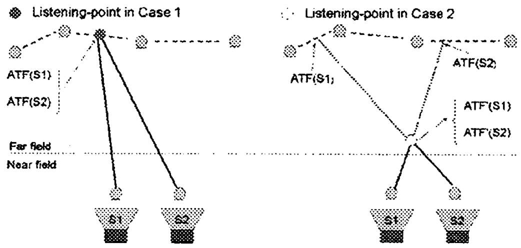

Fig.4: Arbitrary listening-point generation in case 1 and case 2.

4. Experiment

To evaluate the proposed method, an experimental set up is made using 38 microphones on an arc (r=4m) with 20cm distance between each pair, to demonstrate arbitrary listening-point generation, as shown in Fig. 5. Each pair of microphones is connected to a general PC through an amplifier. All PCs are connected to a server PC through a Gigabit Ethernet. The recordings are synchronized by the GPS triggered synchronizer. In addition, there are two sound sources with 1.5m distance that their sounds are recorded by the closely attached microphones without crosstalk.

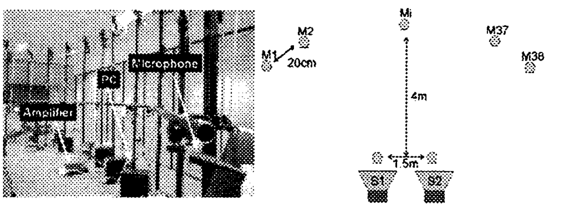

Fig.5: Experimental room and the microphone array configuration.

In the first step, TSP sounds (256 samples, 48000KHz, 16 bits) are broadcasted by speakers in the location of each sound source and the ATFs are generated, then the ATF-pool is generated by synthesizing 20 listening-points between each pair, as it explained in section 2, and section 3.1, respectively. Therefore, ATF-pool contains 1556 listening-points on the arc for two sound sources. Fig. 6 shows examples of ATF for microphone 19 in the array to S1 and S2. The peaks and valleys which are in the ATF are caused by the noise of the PC cluster. In the second step, by using the ATF-pool, arbitrary listening-point generation is performed for a human voice (S1) and background music (S2) that are recorded for a minute. Subjective evaluation of the arbitrary listening-point generation by roaming on the arc array and the listening-point toward the sources is good.

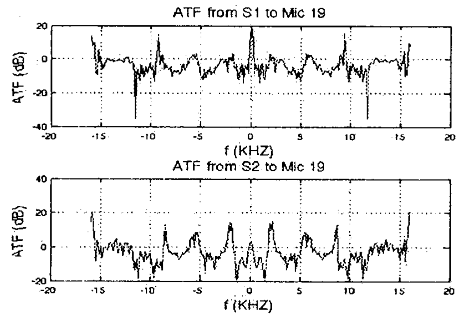

Fig.6: ATF of microphone 19 for S1 and S2.

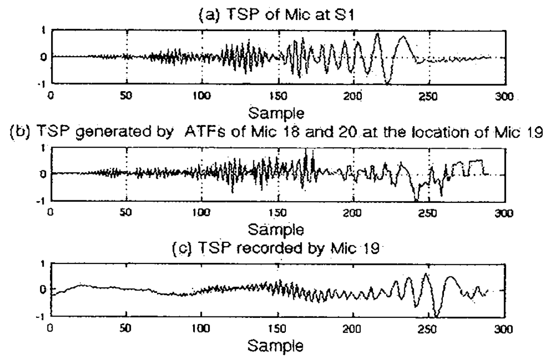

Fig.7: Arbitrary listening-point performance.

For objective evaluation, Fig. 7 shows the generated TSP signal in the location of microphone 19 in comparison with the TSP at speaker and the recorded TSP by microphone 19. The generated sound has less noise (i.e. low frequency noise caused by PC fans) in comparison with sounds captured by the actual microphones in the same location.

Note that this method is not comparable with presented method in [2] due to different arbitrary listening-point generation method, which is done by beamforming techniques in [2].

5. Conclusion

This paper proposed a method to generate arbitrary listening-point using ATFs in microphone arrays. In addition, the representation theory can solve the problem of 3D Image and sound integration, featuring arbitrary viewpoint and listening-point in the same time.

This work has been supported by the SCOPE Fund, Ministry of Internal Affairs and Communication (ref. No.: 041306003).

References

[1] F.L. Wightman, D.J. Kistler, "A Model of HRTFs Based on Principal Component Analysis and Minimum-phase Reconstruction" Journal of the Acoustical Society of America, Vol. 91, No. 3, pp. 1637- 1647, 1992.

[2] M.P. Tehrani, Y. Hirano, T. Fujii, S. Kajita, S. Takeda, M. Tanimoto, K. Mase, "The Sound Wave Ray-Space", IEEE International Conference on Multimedia and Expo, ICME 2005, Amsterdam, The Netherlands, 2005.

[3] P. Na Bangchang, M. P. Tehrani, T. Fujii, M. Tanimoto, "Realtime System of Free Viewpoint Television", The journal of the institute of Image information and Television Engineers (ITE), Vol. 59, No. 8, pp. 63-70, Aug. 2005.