FINAL REVIEW – HRTF PLUGIN "YOUR VOICE^{TM}"

David Perry 460269395 (SID)

Final Review for Digital Audio Systems, DESC9115, 2016

ABSTRACT

The binaural application proposed takes mono audio files and places them within a virtual space dictated by the user. The application processes a recorded voice signal from the user and is then shared as an audio file or within a Multimedia Messaging Service (MMS). The applications primary function is to demonstrate binaural audio quickly with a familiar sound source.

1. INTRODUCTION

Binaural audio and the implementation of Head Related Transfer Functions (HRTF's) is an area experiencing growth and technological development. Several reasons can be proposed for this. Recent innovations in virtual reality (VR) hardware and software have brought VR and augmented reality experiences to smartphones and lounge rooms. Also, the increase in content consumed on headphones due to mobile devices being utilized for content playback and interactivity has very likely contributed.

Introducing smartphone users to the potential of binaural audio would be beneficial to the field and may fuel a curiosity from consumers. By sharing binaural content in the form of voice recorded messages a wider range of people (some of who may be completely unfamiliar) may experience the technology.

2. DESCRIPTION

2.1. Identification

Ones own voice contributes significantly to the perception of the real or virtual environment, and perhaps this characteristic has not been exploited enough in demonstrating the technology. [1] The ease of adjusting the location of the virtual source through a movable 3D model may also encourage experimentation and further interest.

Your Voice^{TM} is a smartphone application that allows users to record their voice (or any other record source) and share it with friends via mobile networks. The rendered output could be incorporated into other platforms such as Instagram, Snapchat or any of the other media sharing applications. Users could add spatial audio to photos and videos, presenting an additional tool for creativity.

2.2. Scope of functionality

The Your Voice^{TM} messages can be sent over a cellular network as a MMS or as data via web service. Obviously messages sent via MMS are limited by data restrictions, while web service messages can be considerably larger. MMS size limits vary from device to carrier, however it was found that data limited to anything less than a 44.1 kHz sampling rate at 16 bit could be considered too poor in tonal quality and spatial information for the purposes of the application. Consequently the duration of the recorded audio was limited to 15 seconds to ensure functionality with MMS.

The spatial processing for Your Voice^{TM} is performed in the time domain, and is based on algorithms proposed in the paper "A structural model for binaural sound synthesis" by Duda et al.

In the paper a useful structural model is proposed that combines an IIR head-shadow model with an FIR pinna-echo model and an FIR shoulder-echo. [3] This model incorporates the much earlier work of Rayleigh, who managed to calculate the equation for the diffraction of a plane wave by a rigid sphere. [4] The signal flow for applying the HRTF's that accounts for the Head Shadow and ITD, Shoulder Echo and Pinna Features are displayed below.

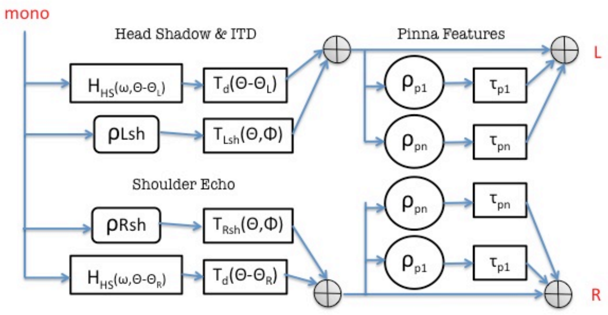

Figure 1. *Structural model for applying HRTF's to a mono channel source. [3]*

Combining an IIR Head-shadow model with an FIR Pinna-echo model and a FIR Shoulder model produces a two-channel audio output with spatial processing.

Inputs required for Your Voice^{TM} to function are: sampling rate (defined), a mono sound source, azimuth angle (Θ) and angle of elevation (Φ).

2.3. Inputs

Sampling frequency and bit depth is assigned as 44,100 kHz, 16 bits respectively. This input value sets the global sampling rate of the application, and is used when creating the audiorecorder object with MATLAB.

The user is prompted to record a short message. Once a recording has been captured (using the smartphones internal microphone or a microphone device) the user can decide at which position the virtual sound source will be created. In MATLAB this information is entered manually, however the data could easily be extracted by selecting a virtual position on a 3-dimensioanl axis displayed within the application. This is demonstrated in a mock up of the applications GUI displayed below in Figure 2.

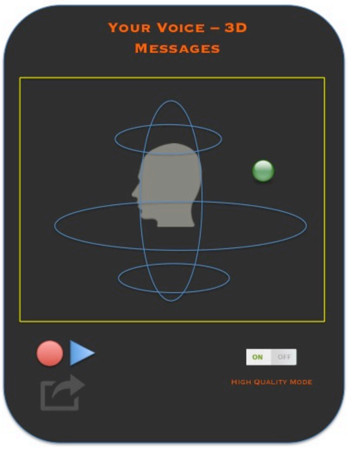

Figure 2. *Users are able to move the 3-dimensional object around the receivers listening position. This would return azimuth (Θ) and elevation (Φ) coordinates used for spatial processing.*

3. EVALUATION

In evaluation test subjects were asked to indicate the azimuth and elevation of an apparent sound source. Two Cartesian planes were provided as displayed in Figure 3. The subjects could only mark the location at a fixed radius, which eliminated any distance factors associated with HRTF's. Three test subjects were used (2M, 1F) none having previous experience with binaural audio.

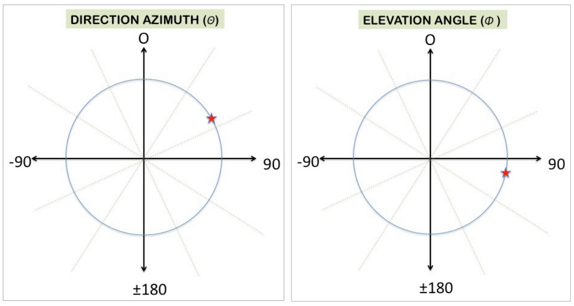

Figure 3. *Users indicate the apparent location of sound source*

Three audio recordings where created for processing and evaluation. The locations of the recordings were interior, exterior and a space with large reflections. Each recording was employed to represent the various locations users may record their voices in everyday use. Background noise was consistent with each environment and most observable in the exterior recording. The evaluation process was applied as followed:

1. Test subjects were played 'interior.wav' with no spatial processing to familiarize the signal.
2. Test subjects were played 'interior.wav' with 90^o azimuth and 90^o elevation.
3. Test subjects indicated on two Cartesian planes the angle and elevation of the apparent sound source.
4. Test subjects were played 'interior.wav' with -60^o azimuth and -30^o elevation.
5. Test subjects indicated on two Cartesian planes the angle and elevation of the apparent sound source.
6. Test subjects were played 'interior.wav' with 120 azimuth and 120 elevation.
7. Test subjects indicated on two Cartesian planes the angle and elevation of the apparent sound source.

The process was repeated with 'exterior.wav' and 'reflections.wav'.

4. RESULTS

Apparent source location results compared with virtual source locations were varied. The most correlated location was observed at 90^o azimuth and 90^o elevation. This was the case across all voice recordings. Interior recording proved to provide the best simulated HRTF, however elevation and azimuth angle greater than 90^o proved almost undetectable. The results for interior are displayed below in Figure 4.

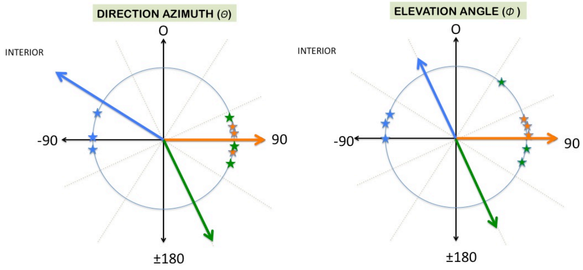

Figure 4. *Coloured arrows represent actual processed location, stars the perceived location. Interior recording demonstrated best sound localisation on the direction azimuth. Localisation of elevation was poor.*

5. LIMITATIONS AND EVALUATION

The application provides fair representation of a defined spatial position given by the user, however elevation and azimuth angles greater that 90 or less than -90 are poorly localized. Further efforts must be made to improve this process before the application is released.

Additionally, as with many binaural processes, one major limitation of the Your Voice^{TM} application is the fact that we rely on dynamic cues for sound localisation. To combat this many developers are incorporating head tracking into applications. The incorporation of head tracking to Your Voice^{TM} would improve the spatial localisation through dynamic cues.

6. DISCUSSION & CONCLUSION

Ideally, the YourVoice^{TM} could be extended to include some form of gameplay. At this stage Your Voice^{TM} is simply an audio messaging application demonstrating binaural audio. To achieve popularity it needs additional functionality, such as gameplay or interactivity. Perhaps users could match each other's source location within a time limit or use head tracking to locate each other in a virtual space. A similar application may allow users to position instruments around the room via the processing of multi-tracked music recordings; a virtual room mixer of sorts. Binaural and 3-dimensional audio is a growth sector within consumer electronics and Your Voice^{TM} aims to introduce more people to technology.

7. ACKNOWLEDGMENTS

Acknowledgment must be made to William L Martens, for whom reading this paper may very difficult given it’s simplification of advanced topics. Considering Bills academic contribution to the field of spatial audio, he should not read this paper directly or it may combust.

8. REFERENCES

[1] Pörschmann, C. (2001). One's own voice in auditory virtual environments. Acta Acustica united with Acustica, 87(3), 378-388.

[2] Zölzer, U., & Arfib, D. (Eds.). (2011). DAFX: digital audio effects (Vol. 1). Wiley.

[3] Brown, C. P., & Duda, R. O. (1998). A structural model for binaural sound synthesis. Speech and Audio Processing, IEEE Transactions on, 6(5), 476-488.

[4] J. W. Strutt (Lord Rayleigh), “On the acoustic shadow of a sphere,” Phil. Trans. R. Soc. London, vol. 203A, pp. 87–97, 1904; The Theory of Sound, 2nd ed. New York: Dover, 1945.

[5] J. P. Blauert, Spatial Hearing, rev. ed. Cambridge, MA: MIT Press, 1997.