

MARRS FOR THE WEB: A MICROPHONE ARRAY RECORDING AND REPRODUCTION SIMULATOR DEVELOPED USING THE WEB AUDIO API

Nikita Goddard and Hyunkook Lee

Applied Psychoacoustics Lab (APL)

University of Huddersfield

h.lee@hud.ac.uk

ABSTRACT

MARRS is an interactive tool that aids recording engineers by establishing the optimal microphone configuration for a desired auditory scene. This makes use of novel psychoacoustic algorithms based on binaural and inter-channel time-level trade-off relationships for both 2 channel and 3 channel microphone array and speaker setups. Previously a mobile app now available on the Android and Apple Store, MARRS is recreated in the web for easier accessibility and further functionality, including the addition of 3 channel microphone array and speaker setups, and the use of the Web Audio API demonstrating phantom image positions of a microphone array across 2 or 3 virtual loudspeakers via binaural rendering.

1. INTRODUCTION

The design and application of microphone arrays for a desired auditory scene can be a difficult and time-consuming task for recording engineers, especially when given limited time for trial and error. This would require manual calculation of the stereo recording angle (SRA) and localisation curves based on the trade-off relationships between inter-channel level difference (ICLD) and interchannel time difference (ICTD) of the corresponding microphone array to achieve a similarity between the source position and the phantom image position. Tools such as William's Curves¹, Wittek's 'Image Assistant'² and Sengpiel's Web Application³ aim to provide a blueprint on the optimal microphone configuration and phantom source prediction, most of which are based previous research on the psychoacoustic principles of auditory localisation. However, these tools hold the following limitations: (i) each tool provides different prediction results due to varying trade-off relationships exhibiting longer or shorter ICTD or ICLD values, (ii) results are only generated for 60° loudspeaker setups, (iii) although SRA and localisation curves are given, there is no visual representation or interaction of the auditory scene. This paper presents the development of a web based tool which descends from the current MARRS, a Microphone Array

Recording and Reproduction Simulator [1] now available on the Apple and Android Store⁴ which already rectifies the limitations of the aforementioned tools. However, further functionality is applied to allow for multichannel formats, improved UI and audio demonstration.

2. PSYCHOACOUSTIC MODEL

The algorithm within MARRS uses a novel psychoacoustic model for phantom image localisation. This is based on the findings of Lee and Rumsey [2] which displayed a linear trading of shift factors between phantom shift regions of 0 to 66.7% (13.3°/0.1ms, 7.8°/dB) and 66.7 and 100% (6.7°/0.1ms, 3.9°/dB) of a ±30° loudspeaker setup. The difference in the upper shift region is derived from the halving of the shift factors as a result of panning uncertainty towards the 30° angle of each speaker.

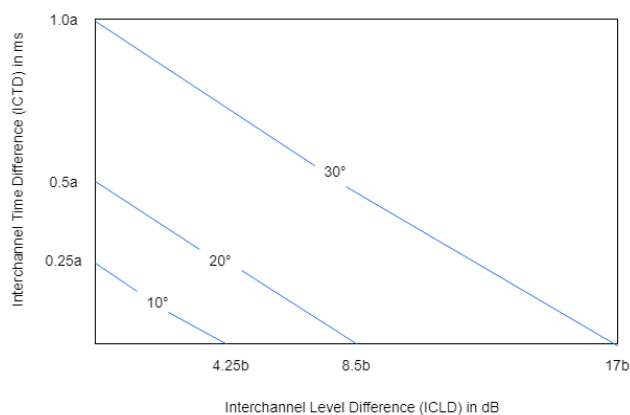


Figure 1: The proposed linear ICTD and ICLD trade-off function. Scale factors along with the values are represented as; $a = (ITD(\theta)/ITD(30))$, $b = (ILD(\theta)/ILD(30))$

For arbitrary speaker setups other than the standard 30° setup, ICTD and ICLD can be multiplied by a scale factor determined by the ILD and ITD of a source angle at 30° and the ILD and ITD of the speaker base angle (see Figure 1). This

¹http://www.mmad.info/MAD/Ch_n_cov.htm

²<https://www.hauptmikrofon.de/stereo-surround/image-assistant>

³<http://www.sengpielaudio.com/HejiaE.htm>

⁴<https://research.hud.ac.uk/institutes-centres/apl/resources/>

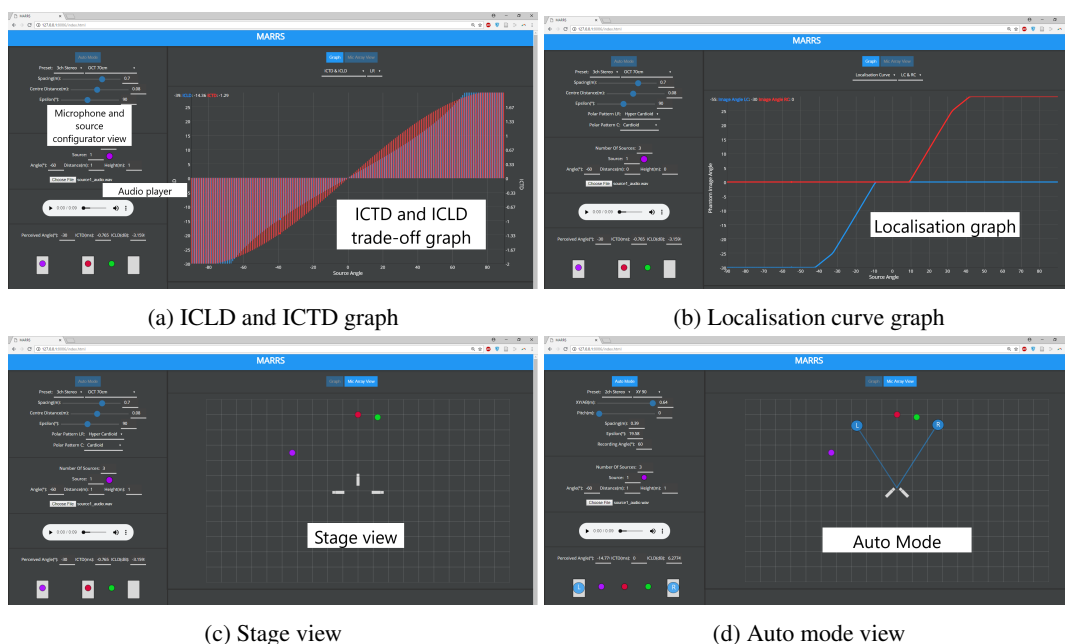


Figure 2: Scenarios and features of the MARRS web app

is derived from the following hypothesis [3]; as the base angle increases, the ICTD and ICLD values of an image shift increases in proportion to the increase of ITD and ILD. This can also be applied to three frontal speaker setups with 5.1 systems which contain two stereophonic segments of 15°, meaning that the scale factor can be determined by the ILD and ITD at 15°.

3. MARRS WEB APP

3.1. Overview

The MARRS web application is accessible on most modern mobile and desktop browsers including Chrome, Firefox and Safari. When loaded, the app displays a single page interface showing the following (see Figure 2); a configuration panel, an audio player view, a loudspeaker view and a main view which can be toggled between the graph view (see Figure 2a and Figure 2b) and the stage view (see Figure 2c). The configuration panel allows full control of all parameters relating to the source position, the amount of sources, the phantom image position and the microphone array, including mic height, mic spacing, subtended angle, polar pattern, etc. Furthermore, microphones array designs can be selected between 3 channel microphone array types in addition to 2 channel mic array types only present in the MARRS mobile app. Below the configuration panel is a loudspeaker view which displays the predicted or desired phantom images between 2 or 3 loudspeakers. On the right of the interface is the main view where the user can switch between the graph view displaying the ICTD and ICLD trading graph and the localisation curves of the corre-

sponding microphone array, and the stage view containing the view of the microphone array and the positions of each source. The ‘auto mode’ button is used to toggle auto mode to allow the user to configure the desired stereo width which will calculate optimal microphone positions (see Figure 2d).

The main feature that differentiates this app from the mobile app and pre-existing tools is the demonstration of the predicted phantom images through virtual loudspeakers, facilitated via the Web Audio API. Further information is explained with the Web Audio section of this paper.

3.2. Architecture

The AngularJS framework was used to develop a single page application which can be accessed on web browsers including Chrome, Firefox and Safari. This allows for easy binding between the functionality of the app, implemented in JavaScript, and the DOM (Document Object Model) elements, the visual UI elements generated via HTML (Hypertext Markup Language) and the Canvas API. NodeJS, a runtime package environment, was used to manage JavaScript files as modules which held specific functions of the app. This meant that development was simplistic and less crowded as scripts can be read into each other.

3.3. Web Audio API

The Web Audio API (WAA) is a realtime advancing audio processing system built into supported modern browsers which is implemented through Javascript. The concept involves a modular structure of audio nodes, each representing an audio process. When the nodes are connected to

each other, the audio chain is augmented and manipulated. For example, the gain of the signal can be changed via the ‘gainNode’. In this instance, a combination of nodes within WAA are used to facilitate virtual loudspeaker rendering so any number of loudspeakers can be represented binaurally. With this approach, users can demonstrate the predicted phantom images between loudspeakers without the need of the corresponding speaker setup, which can be unattainable in terms of space and costs.

For binaural rendering of each loudspeaker, this can be possible through the ‘pannerNode’. However, this does not allow for interchangeable HRTFs. Therefore, a pre-existing library ‘BinauralLib’⁵, created using the WAA, is used to provide the ability to implement HRTFs, with additional features including head rotation and first order ambisonic decoding. The HRTFs are represented in a JSON format which is read into the library, creating a database of positions with their corresponding head related impulse responses (HRIRs). Soon, the library will be adapted to allow for SOFA formats which will become a standard for storing HRTFs. The HRTFs used within the web app are HRTFs derived from a KEMAR dummy head microphone which were limited to a sample length of 512, the same sample length of the HRTFs used within the ‘pannerNode’.

When the user imports a mono audio file to represent a source, a ‘mediaElementSourceNode’ is created. This is connected to the ‘binauraliserNode’ from BinauralLib to create a virtual loudspeaker. Each ‘binauraliserNode’ holds an HRIR relating to the desired position. This is then convolved with the mono input source of the node via the ‘convolverNode’ from the WAA. For a stereo speaker setup, two ‘binauraliserNode’s are created with positions at 30. In order for the phantom images to be accurately represented, the calculated ICTD and ICLD of each phantom source is to be applied to the resulting stereo output of the ‘binauraliserNode’. The stereo signal stream is split via the ‘channelSplitterNode’ into 2 mono signals which can be handled individually. A ‘delayNode’ to apply a time shift in milliseconds and a ‘gainNode’ to apply the resulting gain calculated from the ICLD given in dB are connected. The resulting 2 signals are then outputted via the ‘audioDestinationNode’ (see Figure 3).

4. CONCLUSION

The outcome is a further adaptation of the current MARRS mobile app that is accessible in the web. The highlight of the web app, virtual loudspeaker rendering, demonstrates the high flexibility of the Web Audio API in which nodes can be connected or created for any desired output. The development of the app is still an ongoing process with upcoming updates to the UI showing a side view and a 3D view and the psychoacoustic algorithm of the multichannel

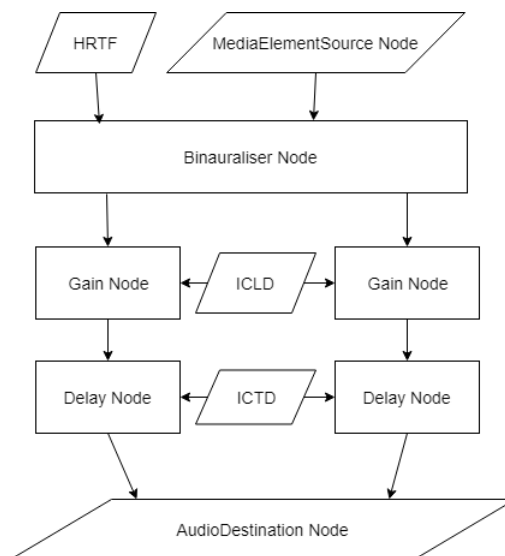


Figure 3: Flowchart of a virtual loudspeaker using nodes from the Web Audio API

microphone array which is currently a research in progress.

5. REFERENCES

- [1] H. Lee et al, “An Interactive and Intelligent Tool for Microphone Array Design,” presented at the 143rd Convention of the *Audio Engineering Society* (2017).
- [2] H. Lee and F. Rumsey, “Level and Time Panning of Phantom Images for Musical Sources,” *J. Audio Eng. Soc.*, vol.61 (12), pp. 753-767 (2013 Dec.).
- [3] H. Lee, “Perceptually motivated amplitude panning (PMAP) for accurate phantom image localisation,” presented at the 142nd Convention of the *Audio Engineering Society* (2017), convention paper 9770.

⁵<https://github.com/Goddarna/BinauralLib>