Analyzing Subjective Selection of HRTFs

Student Researcher: Randale Watson
Johnson C Smith University
r.watson.2010@jcmail.jcsu.edu

Mentor: Dr. Kyla McMullen
Clemson University
kyla@clemson.edu

Abstract* – *Head Related Transfer Function (HRTF) measurements have traditionally been very expensive, technical and time consuming to obtain per individual. However, in recent times there have been techniques implemented and suggested with the goal of making the process more accessible while making it less expensive and efficient. One of those techniques is Subjective Selection, but a major problem is that there is currently no general knowledge of what acoustical cues can be used to predict an HRTF fit for subjects. Subjective Selection data from a study conducted at the University of Michigan was analyzed using data visualization tools in MATLAB computing environment. The data suggested that more selective people do pick HRTFs with similar ITDs. In conclusion Subjective Selection indicates potential to be a reliable method for producing an HRTF fit for an individual.

Background

A *Virtual Audio Environment* is the use of spatial audio software with combination of speaker systems to produce acoustics that are interpreted by the individual hearing these acoustics, to be originating from specific locations of space outside the listener's head giving an audio representation of their spatial location an orientation in the environment (Virtual Audio Systems, 2008).

As the user of the Virtual Audio Environment navigates through the virtual space the software responsible for producing the acoustics will mirror this change by manipulating the respective properties of the audio output to keep the 3D sound experience.

Applications of Virtual Audio Environments:

- Enhanced sound effects in video games.
  - More realistic experience in game play.
- Use in emergency situations to assist visually impaired
  - In emergency evacuation situations where many people are in panic and self-preservation mode.
- Spatial audio interfaces.
- Assisting with navigating an area such as shopping mall or grocery store for the visually impaired.
  - Using spatial sensing devices to relay information to an individual to the isle and cashier queues.

Human Perception of Sounds

The human body like all other natural beings is built to be an extremely efficient and adaptable member in its environment. One of the most amazing concepts about our adaptations to life on earth is the ability of humans to comprehend their wider environment. A great example is our hearing ability to interpret spatial cues using one main organ i.e. the ear (Pamieri, 2011).

A *sound*, in scientific understanding, is the transfer of waves of energy through a medium from the source that compresses and rarefies the materials [vibrates] around it [Fig 1.1]. The process of creating and transferring sound waves are mechanical in nature as sound waves needs a physical medium to pass through.

E.g. Sound through air, in water or in a piece of wood, representing the forms of matter.

Sound waves cannot travel though a vacuum because of its mechanical nature.

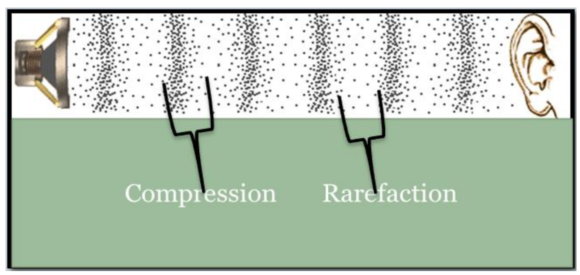

Fig 1.1. Demonstrates the process of sound waves being transfers through air to an individual's ear lobe.

Localization

Humans have been very successful in surviving of this diverse world because of our understanding to take what nature offers us and use it to our benefit. We owe this ability to our sophisticated sensory networks and intellect. To navigate our world we use spatial cues from our senses to understand and navigate our world, the process of doing so is called localization.

Localization is the process of determining where something is positioned, oriented and moving in a space with reference to our relative position is the same space (A. Furmann, 2013).

A basic understanding of how we localize sound is that we can most of the time accurately tell what side of our body is sound closest [Fig 1.2].

Our brain analyzes similarities and difference between the signals from the ears and decodes the signals based on our experience with our world. The brain then gives it best approximation for the source of the sound and its properties that are useful to us.

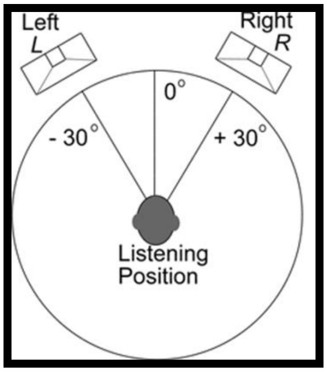

Fig 1.2 Demonstrates sounds coming from different sides of the head and there azimuths.

Head Related Transfer Functions

An HRTF processes and models the change that occurs to a sound in free space to when it reaches the eardrum of both ears, their torso and their head (Xie, 2012).

Duplex theory is a model for estimating an objects spatial location by two binaural cues (Cheng & Wakefield, 1999). We can localize sound from left or right based on phase difference in the sound waves received at both ears (Interaural Time Difference (ITD)) [Fig 1.3]. We also us the relative intensity of sound waves (Interaural Intensity Difference (IID)) difference [Fig 1.4]. This estimation of which direction a sound is coming from is known as the azimuth of the sound.

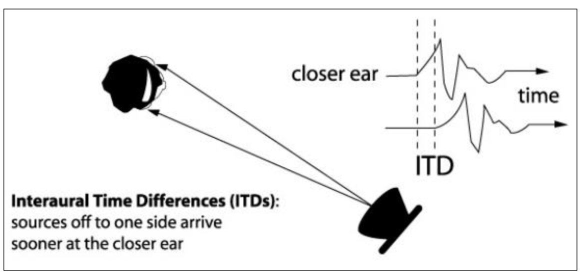

Fig 1.3. Interaural Time Difference: The sound arrives sooner at the ear it is closest to.

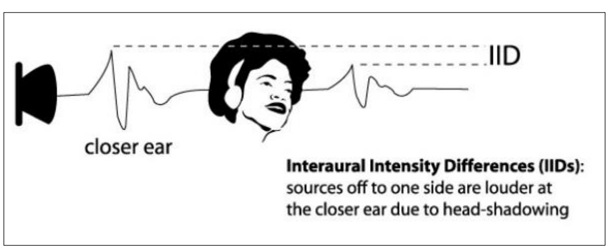

Fig 1.4 Interaural Intensity difference: The sound is louder at the ear that is closer.

Why do we need to measure HRTFs?

People in general differ in their physical make up ever so slightly. These differences are so significant that HRTFs need to be measured on an individual basis. Due to these differences transfer functions are different for each individual.

Differences

- Width and shape of head; a width of the head is directly proportional to the ITD.
- Shape of ear
- Skull differences

Motivation

We know that people can select HRTFs that best suit them based on studies already carried out. At the same time we don't know much about what they are basing their choices on. There are inexpensive alternative methods to the traditional method of measuring HRTFs. There is no measure of what properties are good markers for Subjective Selection. Current methods of creating HRTF profiles are too expensive to allow for wide scale accessibility and production of custom HRTFs. There are databases of measured HRTFs of which many most likely can fit many people.

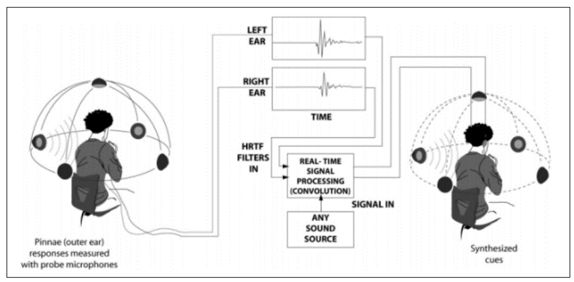

Fig 2.1. One of the modern versions of obtaining an individual's HRTF using a sound dome with loud speakers and microphones in the individual's ears.

We envision the process of selecting an HRTF to be more like shopping for glasses at CVS while the current method is more like getting your glasses made at the ophthalmologist. (Easy Way vs Hard Way)

Research Questions

Interaural Time Difference is the most important cue for us in localizing sounds. Are there underlying pattern in Subjective Selection of HRTF profile. Are there any similarities in the ITD selected by selective listeners compared indiscriminate listeners?

Method

Subjects are presented a list of HRTF profiles from public HRTF database. Each session the participant chooses from the same the same set of HRTF; they are not made aware that they selecting from the same database. The criterions that the participant use to pick a good HRTF include:

- Externalization: If the sound seems to be outside their head.
- Elevation: If they perceive that a sound is above or below their normal plane of hearing.

- Front/Back: If they perceive that the sound coming from in front of them or behind them.

If the participant selects a profile at least two out of three times then the HRTF is deemed selected. There were twenty seven HRTFs used were 13 from the IRCAM (Institut de Recherche et Coordination Acoustique) database, 13 from the CIPIC (Center for Image Processing and Integrated Computing | UC Davis) Database and the KEMAR (Naval Submarine Medical Research Laboratory) dataset

HRTF Preference Interface

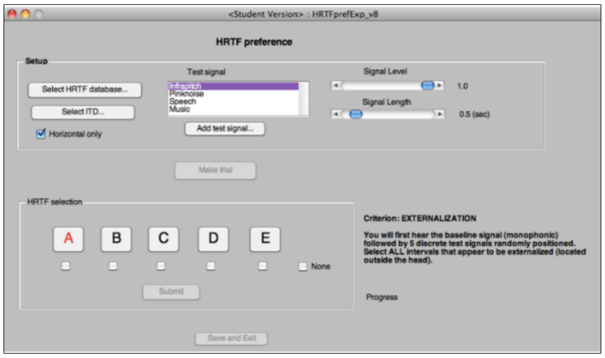

Fig 3.1 Interface for selection tasks. A is the interval currently playing during the externalization stage.

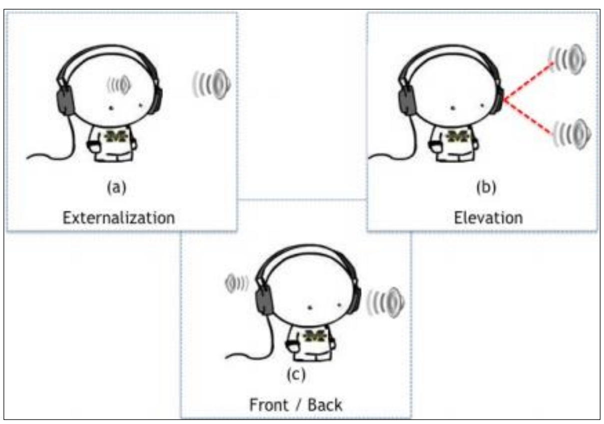

Fig 3.2. Demonstrates different properties of audio cues the participant experiences: externalization quality (a), elevation (b) and front/back (c) distinction

The data collected included the number of hits [selection] for each profile in the list. The properties of the HRTFs selected are then contrasted to observed using data visualizing tools sing MATLAB code.

Terms Used

- Difference score:
  - The sum of the differences between the maximum and minimum time delay measurement among the chosen HRTFs of an individual. This measurement is analyzed for every 30° interval.

- RMSD :
  - Tells how different on average are the much the maximum and minimum time delay point at every 30° interval.

These two methods allow for us to see if the individuals who select fewer ITDs selections have more similarities versus individuals who select several.

Observations

The data was observed for three elevations 0 degrees, 30 degrees and -30 degrees.

Observations at <u>zero</u> degrees of elevation <u>[Fig 4.1.1]</u>

- The difference score does not increase dramatically until a participant selects/prefers seventeen or more ITDs.
- The RMSD does not increase dramatically until a participants selects/prefers eight or more ITDs.
- The rank of selection/preference up to eight from lowest to highest Difference Score are:
  - {(1 : 0),(1 : 0),(1 : 0),(2 : 45), (7 : 55),(4 : 68),(5 : 77) ,(3 : 85),(6 : 91),(8 : 113)}

- The rank of selection/preference up to eight from lowest to highest RMSD are:
  - {(1 : 0),(1 : 0),(1 : 0),(7 : 5.1), (2 : 5.8),(4 : 7.6),(5 : 8.2) ,(3 : 8.9),(6 : 9.2),(8 : 10.7)}

- The three best multiple selection/preference in Difference Scores are at 2, 7 and 4.

- The three best multiple selection/preference in Difference Scores are at 7, 2 and 4.

- The range of selection/preference are from 1 to all 27.

- The average Difference Score was 75.13.

- The average RMSD was 7.38.

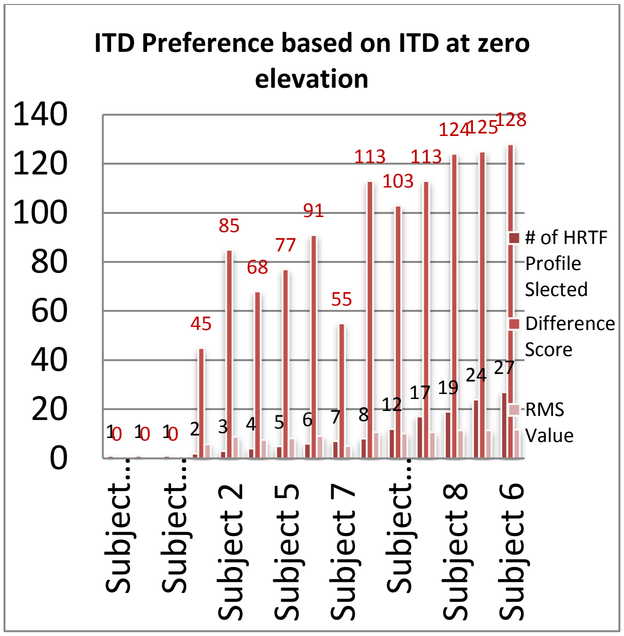

Fig 4.1.1

Observations at *__positive thirty__* *degrees of elevation* [Fig 4.1.2]

- The difference score does not increase dramatically until a participant selects/prefers seventeen or more ITDs.
- The RMSD does not increase dramatically until a participants selects/prefers seventeen or more ITDs.
- The rank of selection/preference up to seventeen from lowest to highest Difference Score are:
  - {(1 : 0),(1 : 0),(1 : 0),(2 : 38), (7 : 43),(4 : 48),(5 : 48) ,(3 : 57),(12 : 63),(8 : 70),(6 : 71),(17 : 74)}

- The rank of selection/preference up to seveteen from lowest to highest RMSD are:
  - {(1 : 0),(1 : 0),(1 : 0),(2 : 3.6), (4 : 3.6),(7 : 3.7),(5 : 4.4) ,(3 : 4.8),(12 : 5.3),(8 : 5.9), (6 : 6.0), (17 : 6.2)}

- The three best multiple selection/preference in Difference Scores are at 2, 7 and 4.

- The three best multiple selection/preference in Difference Scores are at 2, 4 and 7.

- The range of selection/preference are from 1 to all 27.

- The average Difference Score was 50.13.

- The average RMSD was 4.25.

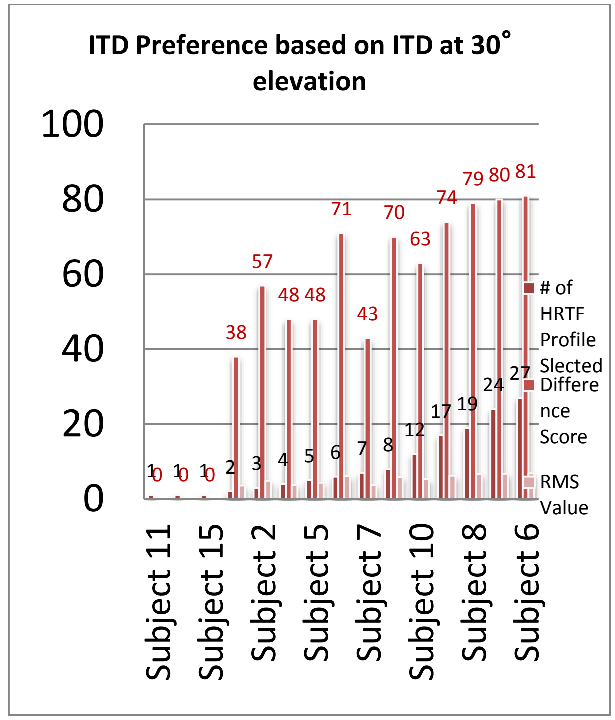

Fig 4.1.2

Observations at *__zero__* *degrees of elevation* [Fig 4.1.3]

- The difference score does not increase dramatically until a participant selects/prefers seventeen or more ITDs.
- The RMSD does not increase dramatically until a participants selects/prefers six or more ITDs.
- The rank of selection/preference up to seventeen from lowest to highest Difference Score are:
  - {(1 : 0),(1 : 0),(1 : 0),(2 : 32), (3 : 47),(4 : 47),(5 : 55) ,(7 : 67),(6 : 69),(12 : 70),(8 : 73),(17 : 83)}

- The rank up 8 selection/preference up to eight from lowest to highest RMSD are:
  - {(1 : 0),(1 : 0),(1 : 0),(2 : 3.2), (3 : 4.1),(4 : 4.1),(5 : 4.9) ,(6 : 6.1)}

- The three best multiple selection/preference in Difference Scores are at 2, 7 and 4.

- The three best multiple selection/preference in Difference Scores are at 7, 2 and 4.

- The range of selection/preference are from 1 to all 27.

- The average Difference Score was 56.6.

- The average RMSD was 4.96.

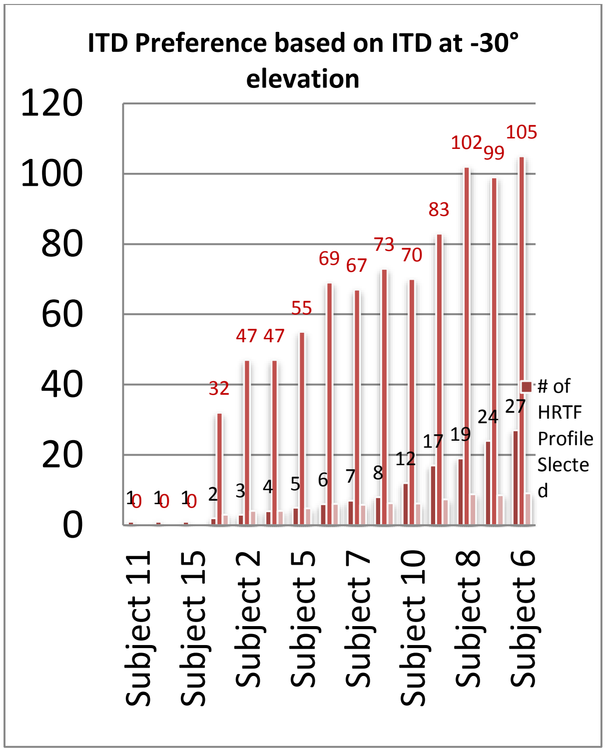

Fig 4.1.3

Data Visualization

The numeric data tells us that in general more selective participants tend to pick more similar HRTF compared to less selective participants based on ITD ques. To verify the implications of the numeric data, data visualization was employed to see how these differences appear graphically. The data was also observed visually by making simultaneous plots on the same graph of all the ITDs selected by participants (subjects).

The graphs plotted have two axes the horizontal axis represents the azimuth that sound was created from while the horizontal axis represents the ITD between the right and left ears.
The straight line from left to right in each ITD plot represents the software keeping continuity from -180° to 150° in azimuth.

Two participants' selection patterns were selected to as data visualization candidates. Subject 7 was chosen to represent more selective persons as this participant selected 7 HRTFs for each of three elevations that were used in the experiment. Subject 1 was chosen to represent less selective people as this participant selected 17 HRTFs for each of the three elevations that were used in the experiment.

Comparisons at zero degree elevation

There are very few noticeable deviations in ITDs for HRTFs selected by Subject 7 [Fig 4.3.1]. There is also consistent pattern in where the peak and troughs occur for each of the plots. In comparison the plots for Subject 1 are more sporadic in from with more deviations and now clear pattern of where peaks and troughs occur for each plot [Fig 4.3.1].

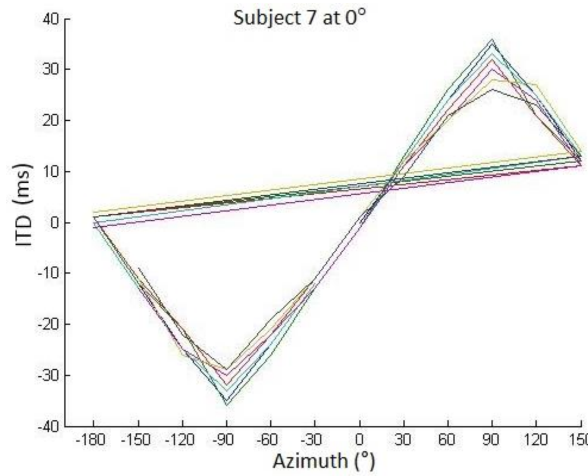

Fig 4.3.1. **Subject 7,** # HRTF picked: 7, Difference score: 55, RMS: 5.0744

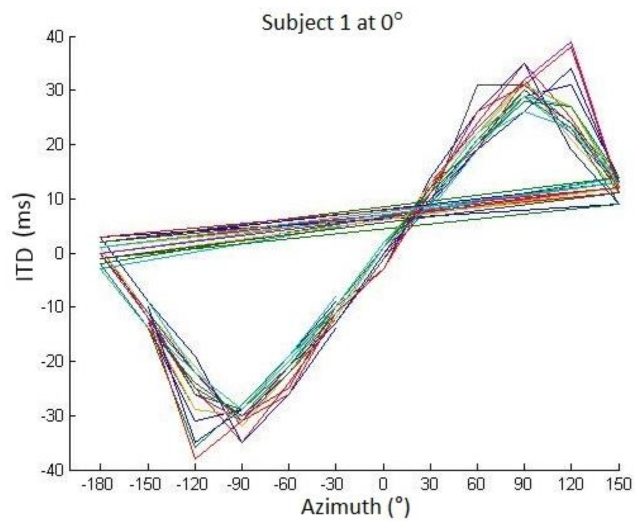

Fig 4.3.2. **Subject 1,** # HRTF picked: 17, Difference score: 113, RMS: 10.6966

Comparisons at thirty degrees elevation

The same observations at zero degree elevations remain true when the two participants are compared at thirty degrees elevations. There are very few noticeable deviations in ITDs for HRTFs selected by Subject 7 [Fig 4.3.3]. There is also consistent pattern in where the peak and troughs occur for each of the plots. In comparison the plots for Subject 1 are slightly more sporadic in from with more deviations and now clear pattern of where peaks and troughs occur for each plot yet is more uniform compares to the observations at zero elevations [Fig 4.3.4].

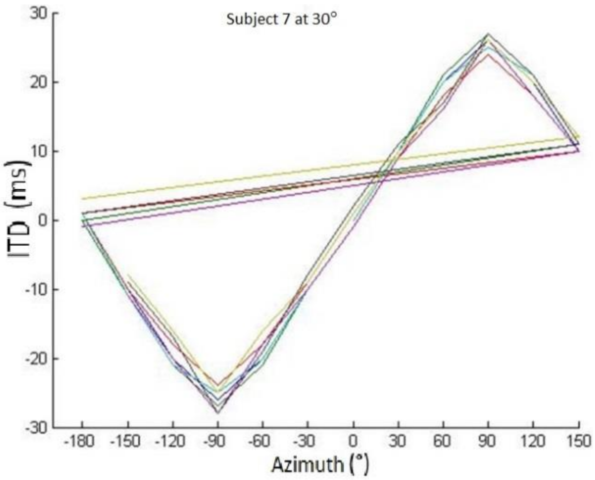

Fig 4.3.3. **Subject 7,** # HRTF picked: 7, Difference score: 43, RMS: 3.7081

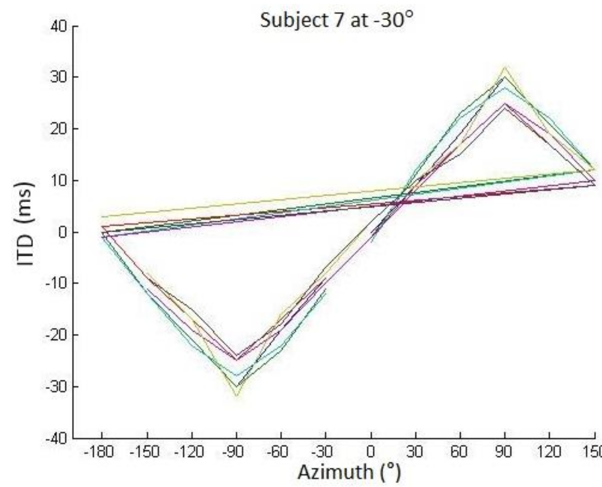

Fig 4.3.5. **Subject 7,** # HRTF picked: 7, Difference score: 67, RMS: 5.8666

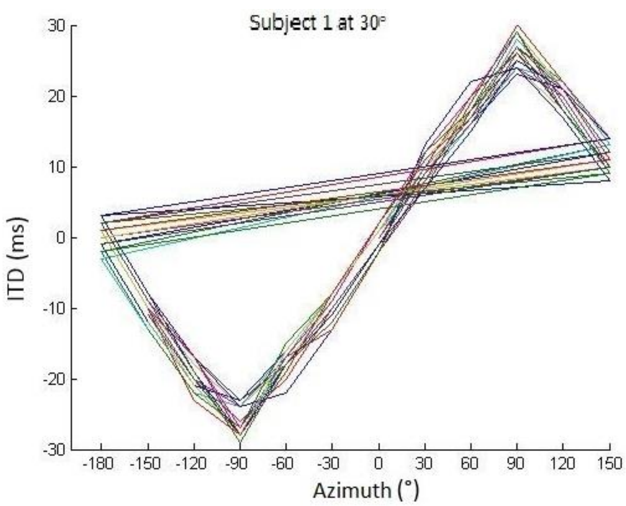

Fig 4.3.4. **Subject 1** # HRTF picked: 17 Difference score: 74, RMS: 6.2183

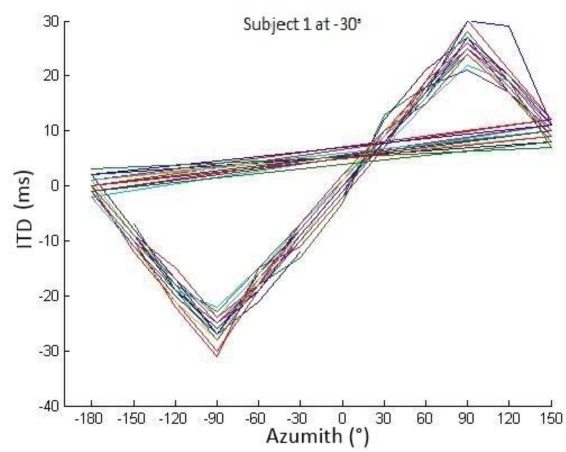

Fig 4.3.6

Subject 1, # HRTF picked: 17, Difference score: 83, RMS: 7.3655

Comparisons at thirty degrees elevation

The same observations at zero degree of elevation and thirty degrees of elevation remain true when the two participants are compared at negative thirty degrees elevations. There are very few noticeable deviations in ITDs for HRTFs selected by Subject 7 [Fig 4.3.5]. There is also consistent pattern in where the peak and troughs occur for each of the plots. In comparison the plots for Subject 1 are more sporadic in from with more deviations and now clear pattern of where peaks and troughs occur for each plot yet is more uniform compares to the observations at zero elevations zero degrees and thirty degrees[Fig 4.3.6].

Reference Data

For convenience of the reader a table [Table 1] with all the data as it relates to trends in selections is included. This is to allow the reader see from the lenses of the writer(s) of this paper

	Elevation 0		Elevation 30		Elevation -30		Average	
# of HRTF Profile Selected	Difference Score	RMS Value	Difference Score	RMS Value	Difference Score	RMS Value	Difference score	RMS Value
1	0	0	0	0	0	0	0.000	0.000
1	0	0	0	0	0	0	0.000	0.000
1	0	0	0	0	0	0	0.000	0.000
2	45	5.7807	38	3.559	32	3.0277	38.333	4.122
3	85	8.8553	57	4.8391	47	4.0723	63.000	5.922
4	68	7.5609	48	3.6286	47	4.1332	54.333	5.108
5	77	8.1904	48	4.3589	55	4.8563	60.000	5.802
6	91	9.1697	71	6.0484	69	6.1305	77.000	7.116
7	55	5.0744	43	3.7081	67	5.8666	55.000	4.883
8	113	10.7355	70	5.902	73	6.2915	85.333	7.643
12	103	10.1366	63	5.3151	70	6.2048	78.667	7.219
17	113	10.6966	74	6.2183	83	7.3655	90.000	8.093
19	124	11.4164	79	6.6395	102	8.8034	101.667	8.953
24	125	11.4054	80	6.733	99	8.5781	101.333	8.906
27	128	11.6976	81	6.813	105	9.0967	104.667	9.202

Table 1

Conclusion and Future Work

More selective listeners tend to select HRTFs with similar cues in ITD based on the analysis of the data. This suggests that selective listeners are finding specific cues when trying to find a comfortable ITD from a premeasured HRTF. Participants who selected fewer HRTF profiles tend to have a lower Difference scores and RMSD values than those who selected a large number of HRTF profiles suggesting that there are more similarities with fewer selections. At the elevations 30°and -30°the both the Difference Score and RMSD values were generally smaller. That suggest that HRTFs are more similar when observed above or below the normal plane of the head. This observation concurs with the difficulty of reproducing and the illusion of elevation with speakers/headphones without Spatial Audio training for individuals (Parseihian & Katz, 2012).

In the future it would be conclusive if the actual measurement of an individual's HRTF was compared to Subjective Selection results. During experiments take greater attention to things like positions of headphones over the ears of participants. Conduct further analysis of different spectral cues.

References

- Furmann, E. S. (2013). The Effect of Sound Reproduction Method on Performance in. ACTA PHYSICA POLONICA A, 988.
- Cheng, C. I., & Wakefield, G. H. (1999). Introduction to HRTFs. AES convention (p. 231). Ann Arbor, MI: Audio Engineering Society.
- Pamieri, P. (2011). A Morphological theory of Human Hearing. AIP Conference Proceedings, 363-367.
- Parseihian, G., & Katz, B. F. (2012). Rapid head-related transfer function adaptation using a virtual auditory environment. Journal of the Acoustical Society of America, (pp. 2948-2949).
- Virtual Audio Systems. (2008). Presence: Teleoperators & Virtual Environments, 527.
- Xie, B.-S. (2012). Recovery of individual head-related transfer functions. Journal of the Acoustic Society of America, 282.