

HEARING WITHOUT EARS

Ian McKenzie

University of Derby
School of Technology
Markeaton Street, Derby. DE22 3AW
England
I.Mckenziel@derby.ac.uk

Peter Lennox

University of Derby
School of Technology
Markeaton Street, Derby. DE22 3AW
England
p.lennox@derby.ac.uk

Bruce Wiggins

University of Derby
School of Technology
Markeaton Street, Derby. DE22 3AW
England
b.j.wiggins@derby.ac.uk

ABSTRACT

We report on on-going work investigating the feasibility of using tissue conduction to evince auditory spatial perception. Early results indicate that it is possible to coherently control externalization, range, directionality (including elevation), movement and some sense of spaciousness without presenting acoustic signals to the outer ear.

Signal control techniques so far have utilised discrete signal feeds, stereo and 1st order ambisonic hierarchies. Some deficiencies in frontal externalization have been observed.

We conclude that, whilst the putative components of the head related transfer function are absent, empirical tests indicate that coherent equivalents are perceptually utilisable. Some implications for perceptual theory and technological implementations are discussed along with potential practical applications and future lines of enquiry.

1. INTRODUCTION

Stimulating auditory perception using mechanical transduction via tissue conduction, bypassing parts of the peripheral hearing system is a known technique. In the 16th century Girolamo Capivaccio employed such by holding a metal rod against the teeth to assess ear pathology [1]. Beethoven produced some of his most well-known pieces at a time when his hearing had completely diminished; he reportedly used a wooden rod, one end held between his teeth and the other resting on the piano enabling him to hear his work [2].

The Fonifero and the Audiphone were early devices that had been developed by the late 1800's, both made use of sound transmission by tissue conduction benefiting the user by delivering an audible experience via alternative pathways. Several binaural tissue conduction devices are available today; the Aftershokz Sportz 2 BC Headphones are marketed for cyclists and the iCharge 4GB Swim Bone Headphones for swimmers. Both allow the user to listen to personal audio while leaving the ears open to the environment, in the case of the cyclist there are obvious safety benefits with this approach.

There are several transmission pathways often collectively referred to as bone conduction transmission pathways (though soft tissues and cerebrospinal fluid feature significantly in these). Four pathways have been identified in numerous research studies as primary bone conduction pathways [3].

- 1) Inertial movement of the ossicular bones relative to the skull at low frequencies
- 2) Distortion of the temporal bone and cochlear shell at high frequencies
- 3) Osseo-tympanic transmission of sound radiated from the walls of an occluded ear canal
- 4) Sound conduction via fluid pathways connecting the cochlea to the brain cerebrospinal fluid

“The resultant sound level at the cochlea is a frequency-dependent vector sum of the contributions from each of these transmission mechanisms.” [3].

Using tissue conduction to evince auditory *spatial* perception is less completely explored. Stanley and Walker [4] in their opening statement discuss prevailing scepticism on the topic; they go on to discuss experiments that indicate that, using inter-signal variations at two transducers contacting the mastoid, lateralisation performance equivalent to that in binaural lateralisation experiments is feasible. Similarly, MacDonald et al. [5] using similar apparatus but alternative locations showed lateralisation performance that was almost identical to those utilising stereo headphones; the transducers were in contact with the condyle just in front of the participant's ears.

Having ascertained that coherent control of lateralisation in auditory perception is feasible via tissue conduction, a more complete range of auditory spatial experience would include the localization of sources and features within a sphere surrounding the perceiver, featuring elevation, front/back discrimination, externalization, range perception, movement perception including auditory looming [6]. It might also include ‘background’ attributes such as spaciousness [7], enclosedness, shape of room, surface textures, and ‘clutter’.

Some of these attributes *may* rely on signal qualities that are insufficiently robust to survive transmission via the multiple paths outlined above; consequently, it is not clear what useful spatial information can be presented to auditory perception in this way. The question which interests us, then, is *whether* and to *what extent* these other spatial attributes may be preserved and coherently controlled.

2. TASK & APPARATUS

Whilst high quality air conducted binaural presentations can supply plausible 3-dimensionality, these rely on coherent control of HRTFs via a one to one mapping of transducer to ear, especially in respect of pinna-encoding. Most, if not all previous research has been conducted using monaural or binaural presentation via tissue conduction with the effect of stimulation at numerous locations on the head documented [8][9]. Some improvement in binaural performance has been realized using generalized Bone Adjustment Functions (BAF) [10] however, due to individuality causing variability in the BAF's it was suggested BAF's may have to be measured for each person. Consequently, a multiple transducer array was selected for initial experimentation investigating the feasibility of controlling azimuth and elevation localization. For simplification, our early experiments feature five transducers which may afford some control of front-back, left-right and up localization using methods more akin to speaker reproduction accepting that the one to one mapping is not the case when using multiple tissue transducers.



Figure 1: Five transducer tissue conduction array showing locations on polystyrene head.

Five tissue conduction transducers were held in place by a flexible plastic framework, a moveable mounting bracket afforded each of the transducers local adjustment. Contact with the head in the desired locations was made via a hard plastic medium and each transducer had its own signal feed.

2.1. Transducers

Five Dayton Audio BCT-1 tactile transducers were used in the device with the following reported specifications:

- Weight. 9 grams each
- Power (RMS). 1 Watt
- Impedance. 8 Ohms
- Frequency Response. 300 – 19,000 Hz
- Physical Size. L/21.6mm, W/14.5mm, H/7.9mm

2.2. Framework and contact force

The framework for the array was made up of flexible plastic headbands connected together. Using average head diameters for Male - 18cm and Female - 17cm [11][1], the contact force exerted by the framework was between 300g (2.49N) to 350g (3.4N). Applications of greater levels of force are reported to improve transmission gain although forces exceeding 5.9N may cause physical discomfort [1]. ANSI standards recommend a force between 4.9N and 5.9N to be used with bone conduction transducers during clinical testing; Bekesy [12] however, found 250g (2.45N) of force sufficient to transmit signals without significant loss.

The size of the contact area with the head also has an effect on transmission with improvements of up to 30dB reported when comparing areas varying from 16mm² to 53mm² [1]. This has been seen to vary with frequency showing improved hearing thresholds above 2kHz, more reliable threshold data was observed with a contact area of 10mm² rather than one of 32mm² [1]; Bekesy used a contact area of 5mm² [12].

The framework provided a suitable force for transmission but it was found that a contact area of 5mm² caused discomfort after a short period time wearing the headset. A contact area of 25mm² provided more comfort but produced significantly more airborne sound. The larger area in conjunction with the type of framework mounting also resulted in poor surface contact of some of the transducers across a range of head size and shape. Several materials differing in size and shape were tried until a 16mm hard plastic semi-circular bead was selected as a contact medium. The semi-circular shape facilitated a more uniform contact across varying head size and shape whilst providing an acceptable level of comfort and sufficient signal transmission; the airborne noise levels produced by the plastic bead were acceptable.

2.3. Locations

McBride et al. [9] conducted a study comparing the sensitivity of the skull with regards to the detectability of signals delivered as a vibrational stimulus via tissue conduction transducers at eleven different locations. Their results were used as a guide to the transducer placements for the tissue conduction array. The consistently best performing location was found to be the Condyle with the Mastoid ranking second.

After consideration of the eleven locations and data obtained from each, five locations were chosen; the Mastoid on both sides, one inch above the Temple location on both sides and a single point between the Forehead and Vertex; transducers were numbered left Mastoid to right.

Although according to the data [9] examined the condyle gave the best performance, it outperformed the other locations considerably; the chosen locations had performed reasonably well and it was the hope they would complement each other. The decision to use only five locations was partly driven by data and the rest by budget and time constraints; five would be sufficient to evaluate the use of multiple transducers.

2.4. Equipment

The following equipment was used for the setup, calibration and test procedures:

- Mac mini3,1 computer with 2 GHz processor and 2 GB memory running Mac OS X version 10.6.8, Firewire connection 800 Mb/sec
- Reaper v4.32 DAW
- Focusrite Saffire Pro 40 multi-channel audio interface
- 5 x 1W LM386N-3 based amplifiers, set Gain of 50 with input attenuation
- Topward DC power supply TPS-4000
- Tektronix TDS3014B Digital Oscilloscope
- Piezo contact microphone
- Digital scales
- Clamp



Figure 2: Equipment setup for subjective and informal testing

2.5. Calibration

When presented with a white noise signal each transducer output was found to vary slightly and required calibrating with respect to each other. Calibration ensured that each transducer, with equal force applied, would output the same level of signal.

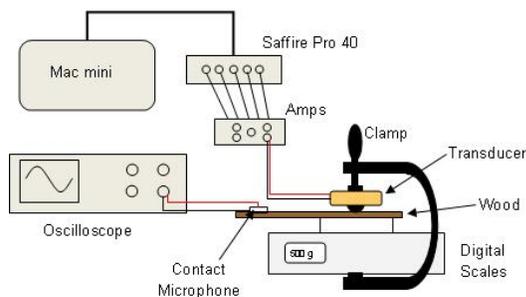


Figure 3: Calibrating individual transducers, each with 300g force applied, amplifier inputs adjusted giving equal output.

The input of each amplifier was adjusted as required to provide equal output at each of the transducers. A white noise signal was then applied in turn, each output recorded and compared. The top signal is the output from each of the transducers, the bottom signal is the reference signal applied to the amplifiers inputs.

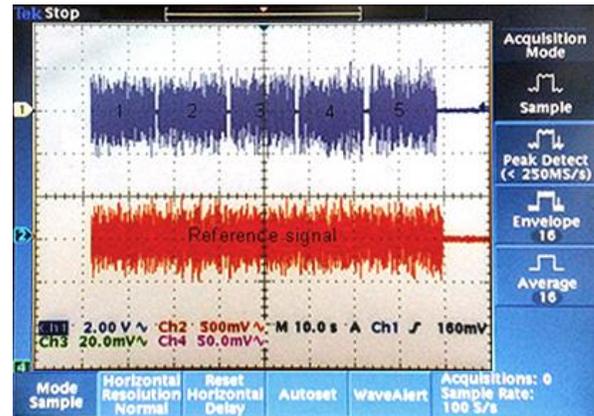


Figure 4: Calibration for each transducer, top signal shows output at each transducer with 300g force applied, bottom signal shows reference signal at amplifier inputs.

3. SUBJECTIVE TESTING

A considerable amount of pilot testing was carried out prior to the subjective testing to establish a short series of tests in areas of particular interest to the project. Lateralization performance via tissue conduction pathways having previously been established [4][5], we were looking for any degree of externalization, elevation or phantom image control and the attributes of any combination of signals providing such.

Subjective tests were carried out with the aim of finding where in relation to the head individuals perceived the test sound to be when stimulated at different locations, singularly or combined. Five subjects, male and female, with ages ranging from 18 to 56 with no known hearing deficits were used during the initial testing.

The tests were carried out with low ambient noise; after calibration the equipment was set up in the same way for each test. Following a brief explanation of the test procedure the participants were seated, the headset put in place and adjusted for location and comfort; ears were not occluded. The test signal, delivered to all transducers equally, was gradually increased to a comfortable level.

3.1. Test One

Pink noise was used as the test signal and presented individually at each transducer in a random order. Participants were asked to identify the area of the head where the sound was perceived to be. The aim of the test was to establish left/right separation and

determine perceived front/back or height, if any, when the signal presented at different locations.

All participants were able to discriminate left and right presented signals but no front to back was observed between the Temple and the Mastoid on the same side, the sound being perceived in or around the ear inside the head. With the test signal additionally presented at location three all participants experienced a degree of height and some front/back along with left/right discrimination.

3.2. Test Two

During pilot testing a degree of externalization had been experienced; the aim of the test was to evaluate any perceived externalization in the participants. A pink noise signal was presented to each participant at the Mastoid and gradually increased in amplitude until the signal was just audible, the same level was then swapped to the Condyle. The test signal, at the same level was then presented at both simultaneously; this was repeated for both left and right sides and comments were invited.

With a lower level of signal presented all participants were able to distinguish a difference in characteristics of the signal between Mastoid and Condyle but the sound remained in or around the ear inside the head. With the signal presented at the Mastoid and Condyle simultaneously all participants experienced some degree of externalization with the sound moving outside the ear.

3.3. Test Three

Signal delay test, the aim of the test was to observe the effect of presenting a delayed signal to one mastoid and the same unmodified signal to the other; a solo singing voice was used as the test signal. The test signal was presented to both Mastoids at an audible level and delayed by 0.65ms at each Mastoid, after a short period one of the delays was removed and the effects via comments observed.

All participants heard the sound initially in the centre of the head, when the delay was removed from one Mastoid it was noted the sound was perceived to move towards that side. The test was repeated several times and the results were the same each time.

The same test was repeated with the addition of location three, the same test signal was presented at all three locations with a 0.65ms second delay. After a short period the delay was removed from one of the locations and comments invited.

All participants experienced signal movement and some phantom imagery; when delay was removed at location three increased height perception was experienced. When delay was removed from either Mastoid the test signal was perceived as externalised to some degree and presenting as between the Mastoid with delay removed and location three.

4. INFORMAL TESTING

Pilot testing had provided areas of interest and a small amount of subjective testing had been in agreement, with very little signal manipulation a degree of height, externalisation and

phantom imagery had been experienced. The following tests were carried out in an informal setting allowing for many different approaches with constant feedback from the participants. The aim of the informal testing was to evaluate how the sounds presented may be manipulated and what effects this might reveal. Test sounds were constructed in the Reaper digital audio workstation software (<http://reaper.fm>) taking unmodified recordings and manipulating the signals with a range of plugins, the same unmodified recordings were also presented Ambisonically over the tissue conduction array.

Test signals:

- Various music pieces in modified stereo format
- Barbershop Quartet - individual modified stems; Bass, Baritone, Tenor and Lead
- 1st order Ambisonic recording of a Motorbike
- A mono recording of a Chainsaw in a forest.

Signal Manipulation:

- Amplitude
- Delay
- Filtering – Highpass, Lowpass, Bandpass
- Phase reversal
- Modified 1st Order Ambisonic decoding using WigWare VST Plugins
- Reverb – FX Plugins and constructed first and late reflections

The range of signal manipulations were chosen as tissue conduction is different from headphone listening as there is not a 1-1 mapping of transducer to ear. In many ways, this makes tissue conduction listening and particularly transducers equidistant from the ears, more akin to loudspeaker listening with respect to phantom image construction and spatial rendition.

Ambisonics is included as a convenient method to manipulate the presented audio as it possesses many useful attributes. As a speaker agnostic with-height system, Ambisonically panned audio can be presented over a multiplicity of transducer arrangements through the use of custom decoders at any azimuth or elevation angle using freely available software tools such as WigWare [13]. All reproduced directions are treated equally, in that preference in the performance of the system isn't given to transducer locations. Ambisonics has also been shown to produce coherent binaural auditory cues for a centrally seated listener [14] and compare favourably to simple pair-wise panning using loudspeaker reproduction [15]. Three dimensional recordings are also available [16] allowing for realistic sound fields to be presented over the transducer array. For this test, a standard 3D cube decoder was used; the decoded loudspeaker signals were patched to combinations of transducers (figure 5) and processed to elicit a comparable directional response to the speaker positions expected by the decoder.

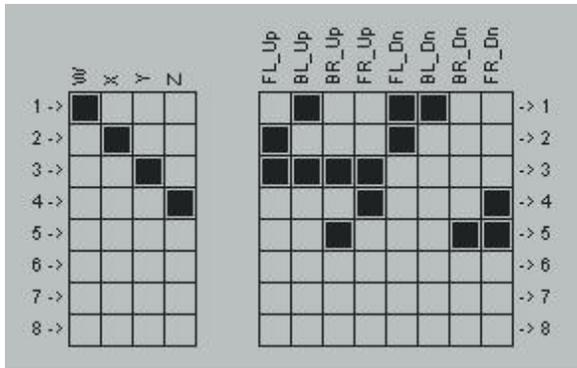


Figure 5: Transducer patch to expected speaker positions, the patching was empirically derived producing a best fit solution.

It was observed during informal testing that there was a settling in period for the participants, when presented with music through the array it took a little while for their hearing to adjust to how the sound was being presented. Many of the participants had not experienced tissue conduction sound before and clearly seemed a little confused by the alternative pathways in use; after a short period they were able to make sense of stereo separation and appreciate a degree of externalization.

After the participants had settled in, the signals were manipulated by adjusting the pre-set delay times, amplitude and filter bands; changes were made and comments invited as the process continued to achieve the best perceived sound for the participant. With the changes made any altered values were noted and relevant adjustments were made to the Barbershop Quartet. An unmodified version of the Quartet was presented first followed by the personalised version and comments were invited about the perceived changes. The Quartet was then presented via first order ambisonic decoding and comments invited.

An ambisonic recording of a motorbike [16] was displayed to the participants several times and comments invited each time. The recording of the chainsaw was used as it presented a clear image allowing the participants to move the image around their heads using the Ambisonic panners; they were asked to evaluate the perceived location against the panner location.

5. DISCUSSION

Pilot and subjective testing provided three areas of interest and sufficient feedback to enable the design of the informal tests. Some degree of height, externalization and phantom image control were the areas we wanted to explore and the informal setting yielded positive feedback.

Amplitude panning via tissue conduction is able to provide a similar lateralisation as to that experienced with headphones [4] [5]. When presented via multiple locations, amplitude by itself was able to produce limited image movement and a small degree of externalization. When amplitude and delay was used to modify signals presented via multiple locations a greater level of image control was realised.

Good separation with a widened image was achieved with the modified stereo presentation, some degree of externalization was experienced and in some cases height and range perception

comments were made. The manipulation of the Quartet stems provided the participants with four spatially discernible voices and information about the type of space they may be performing in; the latter was achieved with reverb plugins and manufactured early and late reflections

The use of filtering was experimented with whilst trying to replicate acoustic head-shadow, normalising for power loss and adjusting delay times yielded inconclusive results at this time; further investigation regarding their use and some creatively designed formal testing may provide us with useful data. The application of filters was used with the modified Quartet feeds to help with image control and externalization gaining positive results.

Quite surprisingly the ambisonic set-up often provided the most positive feedback; the main reason this was unexpected was that the cubic decoder was not designed for use with the array. The alterations made to the decoder were empirically derived yielding positive results; good image control was experienced in the main although frontal presentation of externalised sound was poor. Elevation panning was possible with fairly smooth control and reasonable height achieved. The recording of the motorbike elicited good range and externalization; many of the participants had previously listened to this recording presented via a 24 channel speaker rig and were surprised with the performance of the headset array.

6. CONCLUSIONS AND FURTHER WORK

The on-going project so far has revealed some interesting possibilities; the ability to evince an enhanced spatial image seems plausible. Having considered the feasibility of azimuth and elevation localization, a degree of externalised phantom image manipulation seems possible, further work is required to evaluate and develop these.

The possibility of sound entering via air conduction pathways that may have coloured the participant's perception was a consideration, although test signals were kept low in amplitude. Stanley and Walker (2006) similarly comment in their lateralization tests that due to sound leakage from the transducers it may not have been conducted exclusively via tissue conduction; encouragingly they had shown previously [17] that plugging the ears made little difference to a previous spatial audio task.

Formal testing conducted was of a qualitative nature as perception of a soundfield and related imagery presented via multiple tissue conduction transducers was unknown. The tests were of a more exploratory nature producing areas of further interest rather than quantifiable data. The authors are currently investigating test methodologies in order to elicit more robust results for quantitative analysis.

The use of Ambisonic presentation provided unexpected results considering the decoder used was not designed for the task. Further work will develop the use of ambisonic presentation; the design of a purpose built decoder and test methodology may help us to understand if the need for individualised BAFs is mitigated and why. Future devices may use of different locations employing more transducers with varying levels of force applied over different surface areas, these may be frequency dependant; further papers will follow with data and analysis for consideration.

Further quantitative investigation examining perceptual attributes that may be realisable via a multi-transducer array could include:

- Externalisation
- Elevation
- Image properties (apparent source width ASW)
- Range perception
- Spaciousness
- Movement
- Multiple sources (ref Cocktail party effect)
- Precedence Effects in azimuth and elevation

Testing methodology, signal manipulation and presentation will be considered; Lindeman et al.'s papers provide very interesting insight and will be a consideration as further work is developed [18]. The following points may have some relevance.

- Investigate perceptual training periods
- Investigate correctly tailored decoders
- Investigate how the sounds are captured for presentation against perception
- Investigate higher order ambisonic encode/decode
- Develop reverb for:
 - simplified room modelling (spaciousness)
 - simplified range manipulations

It is our aim to produce a second, further refined and controllable tissue conduction array building on our findings thus far.

7. REFERENCES

- [1] Henry, P., & Letowski, Tomasz R. (2007) Bone Conduction: Anatomy, Physiology, and Communication. ARL-TR-4138.
- [2] Niemoeller, A F. (1940) Handbook of Hearing Aids. New York: Harvest House.
- [3] Dietz, A.J.; May, B.S.; Knaus, D.A.; Greeley, H.P. (2005) Hearing Protection for Bone-Conducted Sound. In New Directions for Improving Audio Effectiveness (pp. 14-1 – 14-18). Meeting Proceedings RTO-MP-HFM-123, Paper 14. Neuilly-sur-Seine, France: RTO.
- [4] Stanley, R., & Walker, B. N. (2006). Lateralization of sounds using bone-conduction headsets. Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES 2006) (pp. 1571-1575), San Francisco, CA.
- [5] MacDonald, J.A., Henry, P.P. & Letowski, T.R. (2006) Spatial audio through a bone conduction interface. International Journal of Audiology, 2006, 45, pp. 595-599.
- [6] Bach, D.R., Neuhoff, J.G., Perrig, W., & Seifritz, E. (2009) Looming sounds as warning signals: The function of motion cues. International Journal of Psychophysiology 74, pp. 28–33.
- [7] Blauert, J., & Lindemann, W. (1986). Auditory spaciousness—some further psychoacoustic analyses. Journal of the Acoustical Society of America, 80(2), pp. 533–542.
- [8] Stenfelt, S., & Goode, R.L. (2005) Transmission properties of bone conducted sound: Measurements in cadaver heads. Journal of the Acoustical Society of America, 118 (2005), pp. 2373- 2391.
- [9] McBride, M., Letowski, T.R., & Tran, P.K. (2005) Bone Conduction Head Sensitivity Mapping: Bone Vibrator. ARL-TR-3556.
- [10] Stanley, R.M. (2009) Measurement and Validation of Bone-Conduction Adjustment Functions in Virtual 3D Audio Displays. Ph.D. thesis, Georgia Institute of Technology.
- [11] Ching, R.P. (2007) Relationship Between Head Mass and Circumference in Human Adults. Ph.D. thesis, University of Washington, Applied Biomechanics Laboratory, Seattle.
- [12] Bekesy, G. V. (1960) *Experiments in Hearing*. New York: McGraw-Hill.
- [13] Wiggins, B. (2014) 'WigWare' *The Blog of Bruce*. [online] Available at: http://www.brucewiggins.co.uk/?page_id=78
- [14] Wiggins, B. (2004) An Investigation into the Real-time Manipulation and Control of Three-dimensional Sound Fields. Ph.D. thesis, University of Derby, Derby, UK.
- [15] Benjamin, E. M., Lee, R., & Heller, A. J. (2010) Why Ambisonics Does Work, 129th AES Convention, San Francisco, CA, USA.
- [16] Soundfield.com. (2013) SoundField: B-Format. [online] Available at: <http://soundfield.com/downloads/b-format.php>.
- [17] Walker, B. N., & Stanley, R. (2005) Thresholds of audibility for bone-conduction headsets. Proceedings of the Eleventh International Conference on Auditory Display (ICAD2005), Limerick, Ireland (6-10 July) pp. 218-222.
- [18] R. W. Lindeman, H. Noma, and P. Goncalves de Barros, "An empirical study of hear-through augmented reality: Using bone conduction to deliver spatialized audio," presented at the IEEE Virtual Reality, 2008, pp. 35–42.