University of Derby

College of Science Engineering

Headphone Image Distortion, Audibility of Acoustic Polarity, and Preserving the Stereo Art

by

Jonathan James Digby
100559332

September 2021

Coursework 2 Report for
7EJ998 Independent Engineering Scholarship

Submitted in part-fulfilment of the requirements for
MH6AB — Master of Science in Audio Engineering

Table of contents

Abstract 3

1 INTRODUCTION 3

2 LITERATURE REVIEW 4
2.1 AUDIBLE POLARITY . . . 4
2.1.1 Inverted Polarity . . . 4
2.1.2 Absolute Polarity . . . 5
2.1.3 Middle Ear Muscle . . . 6
2.1.4 Neuroscience and Auditory Perception . . . 7
2.2 CROSSTALK . . . 8
2.2.1 Stereophonic Sound . . . 8
2.2.2 Interaural Crosstalk . . . 8
2.2.3 Balance and Image Width . . . 10
2.2.4 Head Diffraction . . . 10
2.2.5 Biphonic Reproduction of Stereophonic Material . . . 11
2.2.6 Headphone Response Curves . . . 12
2.2.7 Amplitude Panning . . . 13
2.2.8 Mid/Side Panning . . . 13
2.2.9 Headphone Image Distortion and Preserving the Stereo Art: Part One . . . 15
2.3 THE NECESSITY OF LISTENING EXPERIMENTS . . . 15

3 METHODOLOGY 17
3.1 EXPERIMENTAL DESIGN AND PLANNING . . . 17
3.1.1 Field Experiment . . . 17
3.1.2 Hypothetico-deductive Method . . . 18
3.1.3 Expected Magnitude of Perceptual Differences . . . 19
3.1.4 Duo-trio Test . . . 19
3.1.5 Paired Comparison Test . . . 20
3.1.6 Number of Stimuli . . . 20
3.1.7 Replications . . . 21
3.1.8 Semantic Perception . . . 22
3.1.9 Questionnaire . . . 22
3.2 SUBJECT SELECTION . . . 23
3.2.1 Sample Size . . . 23
3.2.2 Assessor Types . . . 23
3.3 STIMULI SELECTION AND PROCESSING . . . 24
3.3.1 Digital Audio Workstation . . . 24
3.3.2 REAPER Mix Configuration . . . 24

3.3.3 Synthetic Signals and Programme Excerpts: Audible Polarity Experiment . . . 25
3.3.4 Programme Excerpts: Crosstalk Experiment . . . 27
3.4 CONFIGURATION AND CALIBRATION OF THE EXPERIMENTAL SET-UP . . . 29
3.4.1 Listening Test Software Platforms . . . 29
3.4.2 *webMUSHRA* Configuration . . . 30
3.4.3 Loudness Normalisation . . . 30
3.4.4 Level Calibration Stimulus . . . 30
3.4.5 Left-Right Balance Stimulus . . . 31
3.4.6 Training Stage . . . 31
3.5 PILOT AND MAIN STUDY . . . 32

4 RESULTS **32**
4.1 STATISTICAL ANALYSIS . . . 32
4.1.1 Cluster Analysis . . . 33
4.1.2 Wilcox *R* Package . . . 34
4.1.3 Results Summary – Polarity Experiment Duo-trio Test . . . 34
4.1.4 Results Summary – Crosstalk Experiment Paired Comparison Test 35
4.1.5 Voucher Incentive: Winner's Draw . . . 36
4.1.6 Volume Setting Adjustment . . . 36

5 DISCUSSION **36**
5.1 DISCUSSION — AUDIBLE POLARITY . . . 36
5.1.1 Inferred Audibility of Polarity . . . 36
5.1.2 Follow-up Study — Audible Polarity . . . 37
5.1.3 Polarity Standards . . . 37
5.1.4 Audio Waveform Asymmetry . . . 38
5.2 DISCUSSION — CROSSTALK . . . 39
5.2.1 The "missing 6 dB" . . . 39
5.2.2 Difference of the Difference . . . 39
5.3 DISCUSSION — METHOD . . . 40
5.3.1 Allowable Guesses . . . 40
5.3.2 Time Management . . . 40

6 CONCLUSIONS AND RECOMMENDATIONS **41**
6.1 CONCLUSIONS . . . 41
6.2 RECOMMENDATIONS . . . 42
6.2.1 Follow-up Study — Audible Polarity . . . 42
6.2.2 Follow-up Study — Crosstalk . . . 42
6.2.3 Remote Listening Experiment . . . 42
6.2.4 Duo-trio and Paired Comparison Guesses . . . 42
6.2.5 Harman Target Curves and Re-Introduced Crosstalk . . . 43
6.2.6 Unique Perception . . . 43

References **44**

List of Tables **65**

List of Figures **66**

7 APPENDIX **67**
7.1 Supplementary Tables 67
7.2 Supplementary Figures 79
7.3 Experiment Invitation – Email 98
7.4 Determining Number of Assessors 99
7.5 The *R* Project for Statistical Computing 101
7.5.1 *R* Code – Examples 101
7.5.2 *R* code: Gift Voucher Winners 104
7.5.3 Agresti-Coull Method 104
7.5.4 Computational Details 105
7.6 Auditory Demonstration and Test Signals – Sources 106
7.7 Bregman's ' "Auditory Lake" Analogy' (2005) 108
7.8 Nomenclature 109
7.8.1 Explanation of Terms 109
7.8.2 List of Symbols 113

HTML version of this document available at:
https://digbyphonic.com/research/msc2021/20261019reportMSc-DIGBY-2021.html

Headphone Image Distortion, Audibility of Acoustic Polarity, and Preserving the Stereo Art

Jonathan James Digby
100559332

University of Derby
College of Science & Engineering
Markeaton Street, Derby. DE22 3AW
England
j.digby1@unimail.derby.ac.uk

Abstract

A web-based field experiment is conducted to investigate aspects of headphone sound reproduction. The experiment is split into two parts: the first section investigates the audibility of signal polarity for three stimuli; the second section investigates listener preferences between two methods of adding crosstalk to five excerpts of commercially available stereophonic programme material. Relevant aspects of both experiments are reviewed in the literature. Ninety-six self-selecting subjects, casual and expert assessors, used their own hardware to self-administer the listening experiment. The results indicate strong evidence of a discernible difference of acoustic polarity using headphone reproduction for the study population, with source dependent variation. Preference results for the crosstalk experiment were source dependent, with slight evidence of preference indicated for one crosstalk method in two cases, slight evidence of preference for the alternate crosstalk method in one case, and no preference or discernible difference in two cases. Conclusions and recommendations follow a discussion of the findings.

1 INTRODUCTION

The phenomenon of audible polarity, the discernment of a difference between an audio stimulus and its polarity inverted form, is both recognised in the literature and disregarded. Existing international standards or recommendations consider consistent polarity, from input to output, necessary to retain all of the distinguishing qualities of a signal for natural reproduction.[1,2] A remote field experiment is conducted to determine the discernibility of audible polarity with headphone reproduction for three stimuli by a self-selecting random sample of the population, using the subject's own headphones or earphones. This forms the first section of this study.

The second section investigates listener preference for two methods of adding crosstalk to commercially available stereophonic programme material with headphone reproduction. The spatial presentation of stereophonic material using stereo loudspeakers is a result of the physical interactions that occur in the listening environment. With headphone reproduction the spatial presentation, incorporating aspects of stereo loudspeaker phantom image width, front to back depth and tonal balance, is fundamentally different.[3,4]

It may be concluded from reported interviews,[5] anecdotal evidence, and the literature,[6] that the bulk of commercially available stereo programme material produced between the 1960's and the 1990's was intended for and optimised for stereo loudspeaker reproduction. A previous experiment by the author, unpublished report *Headphone Image Distortion and Preserving the Stereo Art: Part One*,[7] attempted to determine if listeners preferred commercially available stereo programme material in its original form or for two methods of re-introduced crosstalk with headphone reproduction. Adding headphone crosstalk may improve a listeners' enjoyment of the existing stereo art by presenting an appropriate resemblance of stereo image width and balance intended by the original artists, producers, and engineers. The use of two simple, least intrusive crosstalk methods may improve and conform the headphone result whilst attempting to avoid unintended consequences or degradations.[8,9]

2 LITERATURE REVIEW

2.1 AUDIBLE POLARITY

2.1.1 Inverted Polarity

Two listening experiments, from 1959 and 1970, showed with certain pairs of opposite polarity stimuli that one subject could consistently judge a stimulus as higher in pitch (or louder or purer), with another subject judging it lower (or softer or less pure).[10,11] Greiner and Melton reported listening experiments with 39 students, using loudspeakers and headphones, showing that polarity inversion was 'clearly audible in some circumstances, but that in many situations it is not'.[12(p246)]

In the literature acoustic polarity is acknowledged as audible with reproduced speech,[13(p575)] Don Davis further reporting that Heyser and Long stressed the importance of 'total polarity being correct',[14(p15)] and that John Hilliard observed and correcting for polarity in theatre sound systems from 1937.[1]

2.1.2 Absolute Polarity

'Absolute polarity',[13(p576)] 'absolute phase',[2(p88)] or 'true phasing'[15(p163)] is the preservation of the signal phase, within the audio band, from microphone to loudspeaker.

The terminal of a microphone has positive polarity when 'an inward movement of a diaphragm[,] resulting from an increase in external sound pressure (compression) produces a positive instantaneous voltage at that terminal, with respect to the other terminal', and for a loudspeaker positive terminal 'a positive instantaneous voltage at that terminal produces an outward movement of a diaphragm'.[16(p8)] Therefore, absolute phase is preserved for a reproduction system if an inward movement of a microphone diaphragm results in an outward movement of the loudspeaker diaphragm.[17(p5)] See Figure 2.1 from Greiner and Melton.[12]

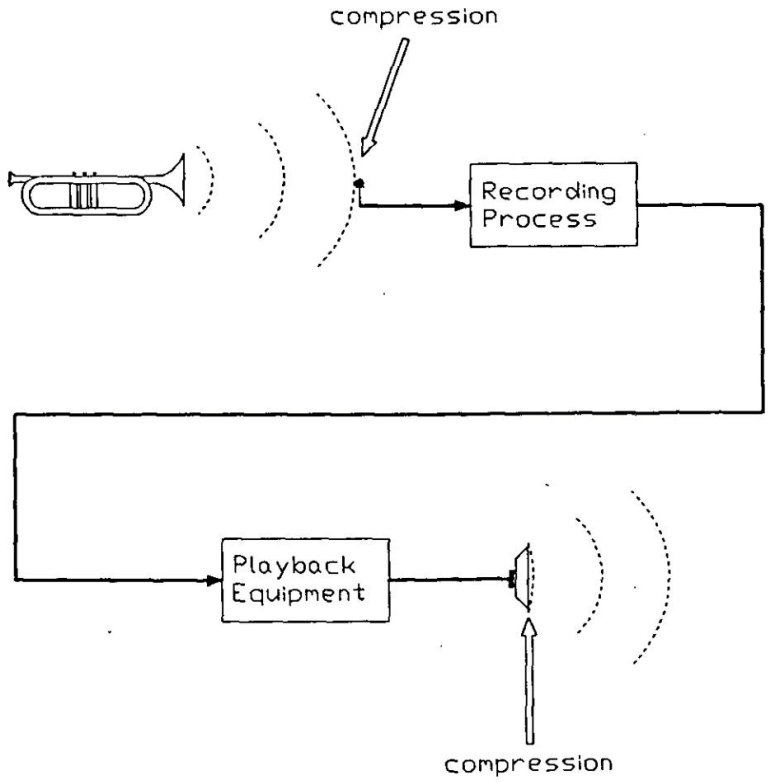

Figure 2.1: 'Record—Playback of acoustic waveform showing that a compression at the microphone should be reproduced as a compression at the loudspeaker'.[12(fig1)] Reproduced from Greiner RA, Melton D; 1994.

Greiner and Melton,[12] Borwick,[2] and Toole,[4] suggest the importance of preserving acoustic polarity for the complete reproduction chain, acknowledging both the subtlety of the effect under many circumstances, and its importance for true fidelity. Johnsen's *Proofs of an Absolute Polarity* contains a chronology, from 1951, of sixty-six printed reports and discussions in support of Absolute Polarity.[18]

In Vanderkooy and Lipshitz's *Polarity and phase standards for analog tape recorders* proposal, the introduction is as follows: 'The neurons in the human ear have a remarkable asymmetry in firing rate depending on the local polarity of the travelling waves on the basilar membrane. It is thus not surprising that the absolute polarity of an audio signal is audible, although the effect is rather subtle'.[19(p2)] Toole states that the basic transduction process in the inner ear performs as a half-wave rectifier,[20] giving 'reason to think that acoustic compressions will be heard as different from acoustic rarefactions'.[4(ch4.8)]

International standards recommending the preservation of polarity from input to output for the film, television, radio and music industry's first appeared in the 1990's.[21–23] The Audio Engineering Society standard AES26, originating in 2001, states 'some audio signals are sufficiently asymmetrical that polarity inversion may be perceptually detected'.[17(p4)]

In contrast to the above, Ohm's acoustic law proposed in 1843, successfully discredited by Seebeck[24] and later revived and elaborated by Helmholtz,[25(p126)] argues that the ear separates complex sound components without regard to phase,[26] Helmholtz stating that the 'ear was phase insensitive'.[27(p48)] The audibility of absolute polarity continues to be questioned in the literature.[28,29(ch3)]

With regard to discernible effects loudspeaker nonlinear distortion must not be discounted. A loudspeaker diaphragm must be capable of equal and opposite movement in the outward and inward direction to faithfully reproduce an input signal. Accurate loudspeaker reproduction may be affected by the diaphragm support, cone stiffness, port design, voice coil design, and so forth.[2(ch2),30–33]

2.1.3 Middle Ear Muscle

The middle-ear-muscle system is found in all classes of vertebrates. Function of the muscles in human hearing remained the subject of speculation until the 20th century, when laboratory experiments on animals and clinical observation made possible the comparative analysis of their physiology.[34]

The middle-ear muscles react involuntarily, with frequency selectivity and attenuation features primed for communication and survival. A person is able to hear other people talking both in the presence of loud background noise, and at the same time as speaking themselves.[35] Fundamental unknowns of the complex neural circuits that control the middle ear muscle reflexes to acoustic and nonacoustic factors continue to exist.[36]

2.1.4 Neuroscience and Auditory Perception

Frequency-following response is recordable from electrodes in the central auditory pathway.[37,38] The phase-locked response 'follows' the waveform of the acoustic stimulus up to a frequency of approximately 1000 Hz.[39,40] Transient peaks in the frequency-following response are shown to match timing and polarity of a stimulus spectrum. Kraus and Nicol state that the fidelity of the brainstem response 'reflects acoustic characteristics of the speech stimulus with remarkable precision in both frequency and time domains',[40(p178)] permitting direct stimulus-to-response correlations. They note that the response might be considered a series of repeated transients. Of interest is their discussion around the signal sampling timescale differences between the left and right hemispheres of the brain, where the left hemisphere samples in the 20–40 ms range, and the right 150–250 ms.

Near instantaneous auditory changes in the non-stationary auditory world[41–43] is an essential ability for survival.[20,44] It may be said it is mainly *difference* detection between the two ears, of timing, phase, amplitude and spectrum, that allows sound source localisation.[45] Heuristic processes are used to separate interleaved and overlapped sounds in natural-world situations based upon grouping the regularities of auditory events from the individuals past experience.[46–48] In this sense the composite mental image formed from the many related but independent perceptions 'has no physical existence…[being]…entirely of the mind'.[49(p257)]

> What you hear is not the air pressure variation in itself but what has drawn your attention in the streams of superimposed air pressure variations at your eardrums. (Linkwitz S, *The Magic in 2-Channel Sound Reproduction - Why is it so Rarely Heard?*[50])

The auditory scene analysis model argues that it is the movement of the listener[51] within the auditory scene that 'provides…the rich information'.[52(p12)] The transfer function of the middle ear autonomically adjusts for social communication and to assess potential threat,[44] where the sensitivity to background noise is attuned to the 'differences between the acoustic patterns registered at each ear'.[53(p10)] This is similarly described by Warren as the *criterion shift rule*, aspects of which 'have been discovered many times…[and] have been given many names, including *adaptation with negative aftereffect, attunement, diabatic effect, persistence effect, central tendency effect, and law*

of contrast'.[54(p582)] For further exploration see Bregman's ' "Auditory Lake" Analogy' in Appendix Section 7.7,[55(p35)] and Boring (1933).[56(ch7)] For interesting parallels with 3D visual perception research see: Mamassian, Knill and Kersten[57]; Zickler, Belhumeur and Kriegman[58]; and Warren.[59(p221)]

2.2 CROSSTALK

2.2.1 Stereophonic Sound

The ANSI / ASA Acoustical Terminology (2020) defines a *stereophonic sound system* as a 'Sound system…used to provide the listener with a sensation of spatial distribution of sound sources'.[60(p42)] The prefix *stereo* in the word stereophonic is borrowed from the Greek *stereos*, meaning "solid", with reference to hardness, solidity, and three-dimensionality. The suffix *phonic* means the science of sound, from the Greek *phone* meaning "sound".

2.2.2 Interaural Crosstalk

An ideally positioned listener, at the point of an equilateral triangle between two loudspeakers may perceive a sound source as a phantom image between the loudspeakers. Stereophonic may, therefore, be described as a perceived "solid" acoustic image in the void between the stereo loudspeakers. Figure 2.2 illustrates the direct acoustic paths from each loudspeaker to the listeners ears.

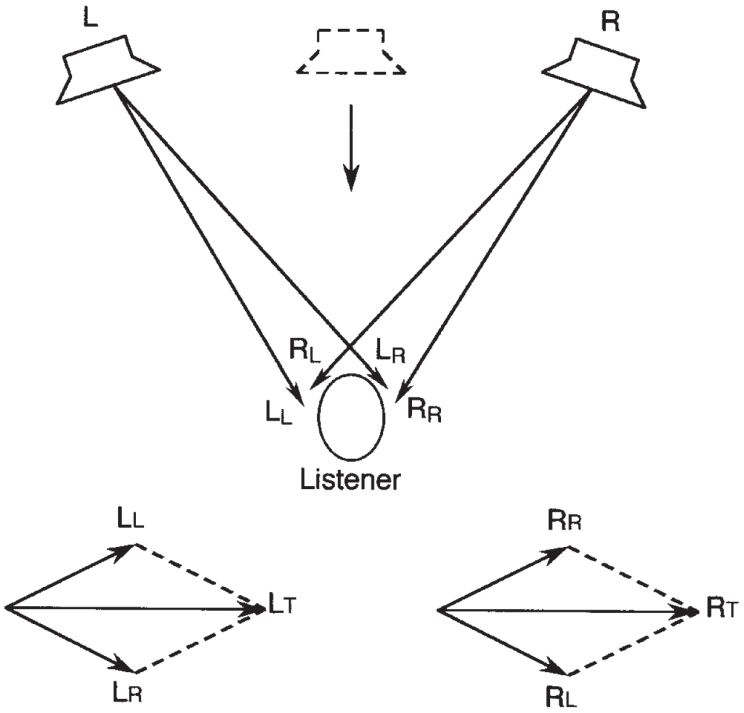

Figure 2.2: Two loudspeakers reproducing the same programme content creating equal net phasors, below 700 Hz, at both ears – reproduced from Eargle J, 2004.[61(figs11–2)] The equal phasors result in a phantom centre image in front of a forward-facing listener. L_L is the signal vector at the Left ear from Left loudspeaker, L_R is the crosstalk signal vector at Left ear from Right loudspeaker. L_T is vector addition of L_L and L_R, a forward phasor for partial contribution to a phantom centre image. Adapted from Bock TM, Keele, Jr. DB (Don); 1986.[62]

The above illustration includes the signals that reach the listeners ears from the loudspeaker at the opposite side of the head, referred to as the *acoustical crosstalk*[3(p16)] or *interaural crosstalk*,[62] abbreviated in this report to *crosstalk*.[6,63]

All virtual image positions between the loudspeakers may be simulated by phasors created from varying the frequency, level, and phase components of the Left and Right loudspeaker signals.[61(p167)] The interaural crosstalk restricts the width of the perceived coherent sound stage, stereo *image width,* to the angular spacing of the stereo loudspeakers.[62]

2.2.3 Balance and Image Width

The ideally positioned listener, in a suitable acoustic environment, may experience a similar distribution of energy, or *balance*, between the Left and Right channels of commercially available stereo programme material as originally determined by the artists, mix engineers and producers.[64] The loudspeaker and room may 'disappear' from the sound scene with accurate two-channel playback in a normal living space,[65(p16)] 'a magic trick for the ears'.[66(p1)]

Unless the listener is correctly positioned on the symmetrical axis between the loudspeakers the phantom centre image and associated spatial relationships are not perceived correctly.[67] For the listener who is no longer along the axis of symmetry the balance is corrupted, through unequal transfer of sound power and asymmetrical directional cues. In turn, the stereo image width is distorted, eventually collapsing to the nearest loudspeaker for the listener sufficiently off-axis. 'Away from the symmetrical axis, in cars and through headphones, we don't hear real stereo; we hear a spatially distorted, but still entertaining, rendering'.[4(ch15.3)]

2.2.4 Head Diffraction

The signal intensity from the left stereo loudspeaker at the left ear remains relatively constant at differing frequencies, whilst signal intensity from the left loudspeaker at the right ear varies with frequency due to added head diffraction and path length difference.[68(p15)] #fig-image6 is reproduced from a paper entitled *Stereophonic Earphones and Binaural Loudspeakers*.[69] This shows a prediction of the sound pressure differences, taken from the studies of Wiener,[70] for the left and right ears across the frequency range for a loudspeaker at an azimuth of 45° in front of the listener. See also Figure 7.21 located in Appendix Section 7.2.

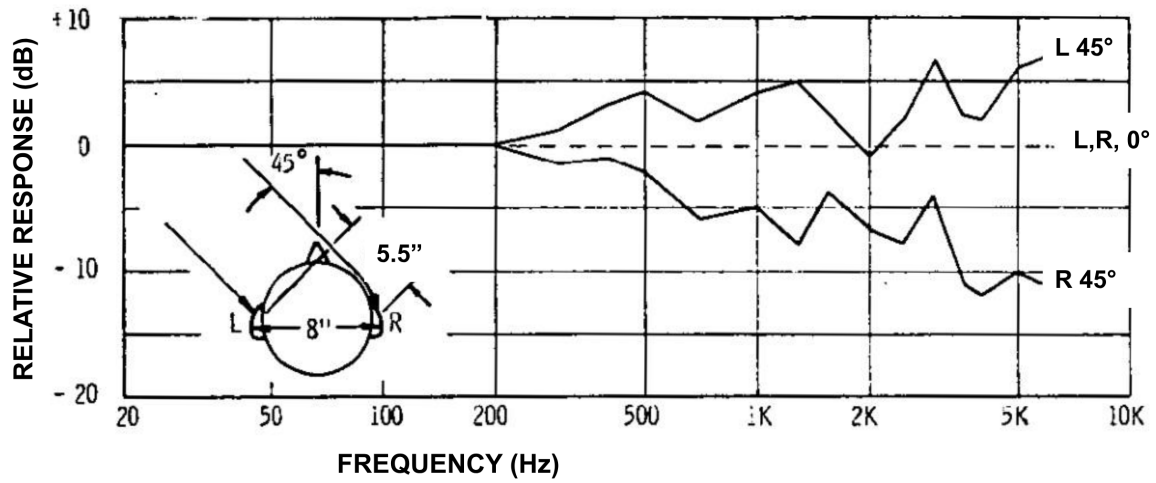

Figure 2.3: Reproduced from Bauer BB, 1961,[69(fig3)] a graph prediction of relative responses at the left (L) and right(R) ears from a loudspeaker positioned at 45° horizontal orientation, derived from Wiener[70]. The "head-on" function (L,R, 0°) is shown as relatively flat. Note between the ear dimension as 8 inches (203 mm), and approximated increased path-length to right ear of 5.5 inches (140 mm). It may be assumed that the common recommended stereo loudspeaker arrangement of ±30° had yet to be realised until a decade or more later.[71(p127),72]

2.2.5 Biphonic Reproduction of Stereophonic Material

The term *biphonic* is used to describe two channel headphone or earphone reproduction,[73(p1.4)] where Left and Right channels are isolated from each other, and it may be said that acoustical crosstalk does not occur.[6(p91)] Without the two indirect acoustic paths to the opposite ear, the crosstalk from the loudspeaker to the ear on the opposite side of the head, the sum energy transfer is altered. Reporting upon the 'highly desirable' finding of substantially preserving the total energy reproduced into a room by loudspeakers from a stereo source, Gerzon states 'if the total stereo energy is preserved, the the [sic] level-balance between different sounds within a stereo mix will be preserved, which is of some artistic importance'.[74(p5)]

With biphonic reproduction it may be stated that a listening room has practically no influence.[23(p5)] This highlights the missing physical interactions within the listening room from loudspeaker on-axis and off-axis reproduction.[65] Additional desirable reflection and diffraction paths associated with the listening room, and the listeners head and torso, the Head Related Transfer Function (HRTF), contribute essential cues to form the spatial impression.[59,75]

Without the spatial cues associated with loudspeakers, biphonic reproduction of conventional stereo or mono sources are perceived “inside the head”,[76(p710)] as originating between the two ears. This is naturally termed In Head Localisation (IHL), also known as “lateralization”.[77] Localisation may be defined as the perceived location of a sound.[78] Two auditory events in the frontal incidence 1° apart in azimuth may be discriminated in binaural hearing, with monaural hearing the ‘localization blur is at least 10 times larger’[79(p2)].

Of interest is a Linkwitz experiment with a listener and two identical loudspeakers in an anechoic room positioned in the standard layout subtending 60°, where, if the listener’s head is not allowed to move and there are no visual cues for loudspeaker location, ‘the listener hears a phantom source inside his head between the ears’.[66(p2)] When the listener is allowed to turn their head, the phantom source is perceived in front at approximately the distance to the invisible loudspeakers.

2.2.6 Headphone Response Curves

In IEC standard 60268-7 *Sound system equipment Part 7: Headphones and earphones* specifies two categories of headphone frequency response calibration, the free-field comparison and the diffuse-field comparison,[80] ‘because no method has yet been developed that is universally applicable’.[81(p25)]

In contrast to the recommended headphone calibration method using an HRTF from a loudspeaker in an anechoic environment (free-field) or a reverberation chamber (diffuse-field),[82] the experimentally derived Harman target curves[83–89] follow this sensible logic: ‘Since stereo recordings are optimized to sound best through loudspeakers in rooms … stereo recordings will sound best when reproduced through headphones that simulate the in-room response of a well-designed loudspeaker system calibrated in a reference listening room’.[90(p24)] See Figure 7.19 located in Appendix Section 7.2.

It should be noted that the preferred room correction for the loudspeakers, setup with an incident *stereo angle* of ±30°, was a smooth 10 dB negative slope from 20 Hz to 20 kHz.[91] This room target curve is in agreement with Toole,[4(ch11),92] and may be seen as a minor modification from the 1974 recommended response curve in Bruel and Kjaer’s *Relevant loudspeaker tests in studios in Hi-Fi dealers’ demo rooms in the home etc. using 1/3 octave, pink-weighted, random noise*.[93(p4)] See Figure 7.20 located in Appendix Section 7.2.

2.2.7 Amplitude Panning

The term ‘panning’ has its origin in the motion picture industry, a ‘panoram’ tripod head for a camera ‘providing for the panoraming movement or horizontal swinging movement of the camera upon the tripod’ (American School of Correspondence and Hulfish, 1911).[94] Audio panning is the positioning of an image source within the stereo sound field. A mono source may be spatially positioned in the phantom image between stereo loudspeakers through interaural arrival-time differences, ITDs,[95] and interaural level differences, both of which are frequency dependent.[79]

Figure 2.4, from Pulkki[96(p458)], illustrates the positioning of a virtual source on the active arc between loudspeaker ‘channel 1’, and loudspeaker ‘channel 2’, using vectors. Altering the ratio of the gains $\mathbf{g_1}$ and $\mathbf{g_2}$ upon unit length vectors $\mathbf{l_1}$ and $\mathbf{l_2}$ combine to produce vector $\mathbf{p}$, which points toward the virtual source.

For this report the amplitude panning method for directly adjusting the Left and Right channel output gains of an input source is denoted by PAN.

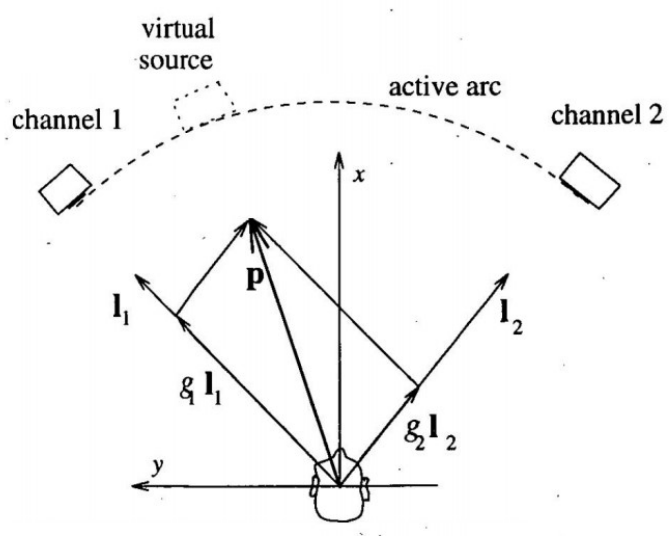

Figure 2.4: Stereophonic amplitude panning formulated with vectors: $\mathbf{p = g_1\, l_1 + g_2\, l_2}$ where $\mathbf{l_1\ l_2}$ are loudspeaker vectors; $\mathbf{g_1\ g_2}$ are the gains or amplitude levels for loudspeaker **channel 1** and **2**. Altering the relative gains alters the virtual source position $\mathbf{p}$ on the active arc. Reproduced from Pulkki V, 1997.[96(fig3)]

2.2.8 Mid/Side Panning

Mid/Side[73(p7.14)] panning uses the “sum” and “difference” signal matrix technique from Alan Dower Blumlein’s patent from 1931, *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems.*[97] Blumlein’s seminal work proposed the placement of two loudspeakers either side of a cinema screen to reproduce the recorded matrix output of the “sum” and “difference” from two

microphones, to 'produce differences of intensity at the loudspeakers so as to give an impression to an observer of oblique sound incidence'.[97(p94)] Incidentally, Blumlein saw 'no reason' for not using the existing "stand-by" channel, the backup safety, for the new second channel.[97(p95)]

$$\text{sum} = L + R$$

$$\text{difference} = L - R$$

Figure 2.5, from Gerzon,[98] illustrates the signal flow for a Left and Right input, an *MS MATRIX* first encodes into *M* (Mid, Mono, or "sum") and *S* (Side or "difference"), the subsequent matrix decodes back to *L*? and *R*? stereo signal outputs. The Width Gain control for the *S'* channel alters the ratio between Mid and Side, thereby adjusting the stereo image width (Schneider, 2012).

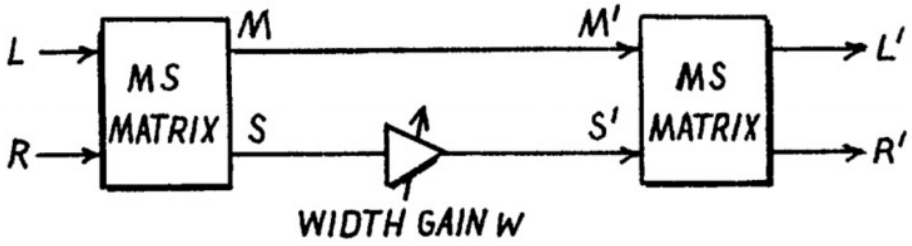

Figure 2.5: Schematic of a width control for amplitude stereophony. The "sum" channels are here denoted as MID signals **M** and **M'**, the "difference" channels denoted as **S** and **S'**. Reproduced from Gerzon MA, 1992.[98(fig2)]

The interchangeable terminology of "sum" and "difference" and Mid/Side refers to the application of Blumlein's matrixing technique for a forward-facing Mid microphone capsule, and Side oriented figure-of-eight microphone capsule.[99(p154)] See Figure 7.28 located in Appendix Section 7.2.

The equations below show the signal flow from Gerzon's Figure 2.5, illustrating the identical operation of both instances of *MS MATRIX*:

$$M = L + RS = L - RM' = MS' = S \times WL' = M' + S'R' = M' - S'$$

For this report the Mid/Side panning method for adjusting the stereo image width of a Left and Right output from a stereo source is denoted by MS.

2.2.9 Headphone Image Distortion and Preserving the Stereo Art: Part One

In an unpublished report by the author,[7] crosstalk was introduced to commercially available stereo programme material using the MS and PAN methods described, at three crosstalk levels. The empirically derived biphonic crosstalk levels, auditioned to naturally characterise the spatial attributes perceived with loudspeaker reproduction,[100] were calibrated by the difference in level between Left and Right output channels from a signal sent to either left or right input channel only, being 6 dB, 8dB, and 10 dB respectively. The goal of the experiment was to determine preference scores between a reference original, the three PAN processed versions (6 dB, 8 dB, 10 dB), and the three MS processed versions. Nine commercially available stereo programme material excerpts were used.

Thirty-one (31) participants completed the remote field experiment. The "Multi Stimulus test with Hidden Reference and Anchor" (MUSHRA) test method was employed,[101] using the web application *webMUSHRA*.[102]

Bias-corrected and accelerated (BCa) bootstrap 95 % confidence interval analysis of the results showed the commercially available stereo programme material in its original Left-Right format was not clearly preferred for the majority of programme excerpts. This corresponded with reported findings and experiments of biphonic reproduction of stereophonic material in the literature.[6,69,103–106]

The amplitude panning (PAN) and sum-and-difference methods (MS) received different preference ratings that was not consistent across all programme excerpts. The three levels of introduced crosstalk, six (6), eight (8) and ten (10) decibels, also received differing preference ratings for a given excerpt, with a general trend toward 8 dB and 10 dB, dependent upon the programme material.

In general, the results were statistically inconclusive, however trends were observed with further investigation recommended.

2.3 THE NECESSITY OF LISTENING EXPERIMENTS

A physical measurement of an audio signal in the acoustic or electrical domains does not explain how the human auditory system interprets the event.[107] Heyser states 'much of the discrepancy between how we measure and how we hear lies in nonlinear effects', stressing the 'importance of attempting to bring objective and subjective audio together'.[30(p674)] With 'the ear as one of our finest acoustical instruments for measuring sound',[108(p534)] each listener's unique perception is necessarily influenced by their experience in the present, and in relation to their past.

> The mind, eager to accept signals from all possible sources, takes its cues from a number of sensory organs … Our perception of signals from all muscles and joints together with signals from our semicircular canals of the inner ear supply the basic reference to the three-dimensional world … Psychologists can prove that the mind is not even capable of simultaneously registering a multitude of divergent perceptions with all their details. (Pickering NC, Baender E, *Two ears in three dimensions*[49(p258)])

In 2010, Evans and Treisman showed mounting evidence that perceptual experiences that appear to be modality-specific, for example solely auditory or solely visual, are influenced by activity from other sensory modalities on both the conscious and the unconscious level, 'in an automatic fashion'[109(p1)] — Figure 2.6 illustrates.

In a 2017 study, subjects[110] were able to reliably categorize speech, music, and human-environmental sounds after as little as twenty-five milliseconds of sound exposure.[111] A section of the Bizley and Cohen review of auditory-object perception research noted that 'attention is not required for a person to detect changes in a stimulus feature', concluding that the 'roles of different neural structures…are not yet fully understood'.[112(p693)] Fundamental questions of perception remain unanswered in 2020 according to *Predictive coding in auditory perception* in the European Journal of Neuroscience.[113]

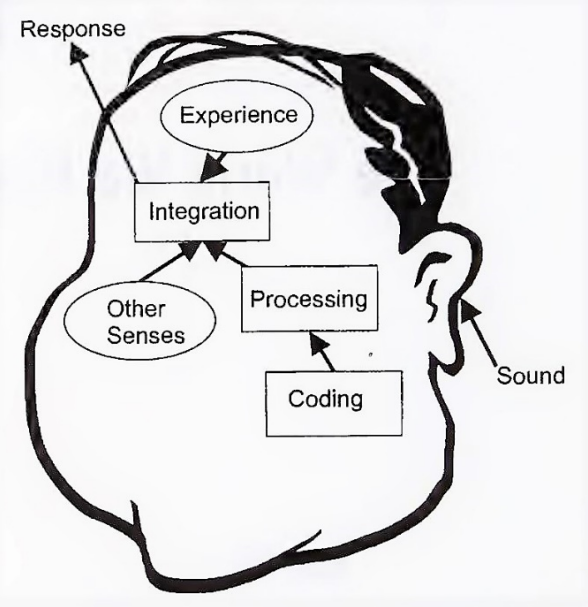

Figure 2.6: A physical sound field stimulates the ear, resulting in neural responses that combine to form perception of the auditory event. The frequency, intensity and time attributes of sound are first coded by the peripheral auditory nervous system. This neural code is processed by higher neural centres to allow the listener to determine the sources of sound. This neural information is integrated with other sensory information, and that of experience, leading to behavioural responses. Reproduced from Yost WA, 2013.[114(fig1.2)]

It remains the case, to date, that an experimenter cannot perceive the subject's auditory event directly, but must instead rely on the subjects feedback in relation to the experimenters' own experience.[63] Listening experiments, therefore, 'remain the final arbiters of sound quality'.[115(p2)]

3 METHODOLOGY

3.1 EXPERIMENTAL DESIGN AND PLANNING

> The chosen design should explore the shadowy regions of present knowledge whose illumination is currently believed to be important to progress. The design is represented by a movable window through which certain aspects of the true state of nature, more or less distorted by noise, may be observed.
>
> — George EP Box, William G Hunter and J Stuart Hunter; 1978.[116(p5)]

3.1.1 Field Experiment

A web-based field experiment was required under the circumstances of the coronavirus disease 19 global pandemic a laboratory experiment being unachievable.[117]

A remote web application allows the listener to perform the experiment in their natural environment, in contrast to a laboratory setting, prioritising the listener as the crucial element of the experimental process.[118] Advantages over a laboratory based study include the subject's ability to perform the experiment at a time and location of their choosing with their own playback equipment,[119] the willingness to participate being an essential element of sensory analysis.[120(p2)] These factors combined may result in a greater number of participants for the duration of the study.[121]

On the contrary, a laboratory setting allows consistency of equipment, environment, ambient noise, system calibration, one-on-one instruction, and supervision. The ITU Recommendation BS.1534-3[101(p6)] suggests a large number of assessors might be needed to achieve sufficient resolution where tight experimental control cannot be achieved.

> With the control of a laboratory experiment … comes an artificiality that raises questions about the generalizability of results. And yet the "naturalness" of a field study … can jeopardize the validity of causal inferences. (Visser PS, Krosnick JA, Lavrakas PJ, *Handbook of Research Methods in Social and Personality Psychology*[122(p223)])

Under these conditions the uncontrollable variables, such as headphone or earphone type and model, experiment environment, and so forth, may be considered aspects of an individual's regular procedure when using headphone or earphone reproduction. If the results of the study allow a hypothesis to be falsified for the sum of this range of different conditions, it may be inferred under these same conditions to the population. More importantly, if the field experiment shows significant effect in the subjects' natural environment,[123] where uncontrollable field factors may be viewed as necessary conditions, resulting inferences may be considered to be all the stronger.[124–126]

An international Standard outlining an approach suitable to the conditions of this experiment was not found.[107]

3.1.2 Hypothetico-deductive Method

The goal of the two experiments, through Popper's[127] hypothetico-deductive method, is to determine at a suitable level of confidence if the following statements may be falsified for headphone or earphone reproduction:

- Subjects do not hear a difference of audible polarity between a stimulus and its inverted polarity form
- Subjects do not perceive a difference between the two crosstalk methods, MS and PAN, with excerpts of commercially available stereo programme material
- Subjects do not prefer either of the two crosstalk methods with excerpts of commercially available stereo programme material

Independent variables include the selected synthetic test stimuli and selected programme materials (genre, variation, etc.). Dependent variables include the physical and response properties of the subject's headphone or earphone, the experiment environment (ambient noise, temperature, etc.), the number of subjects and assessor type, and so forth.[107]

3.1.3 Expected Magnitude of Perceptual Differences

The history of the audible polarity phenomenon suggested the expected magnitudes of the perceptual differences may be small. In the prior crosstalk experiment,[7] differences between objects were observed, as expected, but were not statistically significant, hence the current crosstalk experiment attempts to establish a conclusion.

For declared results of subjective tests to be meaningful care must be taken to reduce differences between the subjects' for their ratings of similar intention. In the prior experiment[7] a bootstrap method was necessary for the statistical analysis due to nonnormal distributions.[128–130]

Lipshitz and Vanderkooy[131] advocate the use of blind, or double-blind, A/B testing, it being considered a robust and simple method suitable to the criteria of a field experiment.[132] The binary choice allows a simultaneous rating of all attributes, and is both statistically and time efficient.[133] With the dependent variables of a field experiment it is deemed inappropriate to compare subjective attributes[115,134,135] beyond a binary decision preference, due to the associated bias effects upon ratings.[136]

3.1.4 Duo-trio Test

For the audible polarity experiment a duo-trio test method is used,[137] an effective method for determining that a perceptible difference exists between a pair of items. Subjects are presented with three items simultaneously, item *A*, item *B*, and a *Reference* item. In each trial an item and its inverted polarity form are randomly assigned to either *A* or *B*, in random order. The *Reference* item is a randomised selection of *A* or *B*. For this experiment the subject is instructed to choose the item *A* or *B* that matches the *Reference* item. The null hypothesis H_0 states the probability of the population making a correct decision when there is *no* perceptible difference between the items is one in two:

$$H_0 \,:\, p = 0.5$$

Stated another way, the null hypothesis corresponds to the equal probability of "success" or "failure" over multiple trials where subjects do not detect a difference in the items.

The alternative hypothesis H_1 states that if there *is* a perceptible difference between the items the population would match the reference with the correct item more than one in two times:

$$H_1 \,:\, p > 0.5$$

The duo-trio test is a forced-choice procedure, where subjects are not allowed the option of reporting “no difference”. The decision was made to not use the word “guess”, or to instruct subjects to guess if unsure, due to a concern of a possible lack of engagement that this may infer. An apparent lack of engagement, in this case, would be undetectable with an unsupervised remote field experiment. The invitation and instructions for the experiment included the following sentence: ‘it is not a test, listening preferences are all that’s required, there are no wrong answers’.

3.1.5 Paired Comparison Test

The stereo crosstalk experiment is a directional two-tailed test of a binomial trial which may be viewed with the following hypotheses:

$$H_1 \,:\, p < 0.5$$

$$H_2 \,:\, p = 0.5$$

$$H_3 \,:\, p > 0.5$$

The null hypothesis, H_2, represents no preference for either item. The alternative hypotheses H_1 H_3 represent a preference in the direction of the associated item, where a choice of item A scores ‘0’ (zero) in the direction of hypothesis H_1, and a choice of item B scores ‘1’ (one) for H_3.[138]

3.1.6 Number of Stimuli

The required number of stimuli for each experiment was guided by the ITU-R BS.1534-3 recommended estimate of ‘1.5 times the number of systems under test, subject to a minimum value of 5 excerpts’.[101(p13)] Seven crosstalk samples were sought for the pilot study, and five polarity samples, surplus to the expected final requirement.

The desired total length of time required to complete the experiment is between fifteen and twenty minutes.[107(p302)] This allows participants to complete the task without tiring. A reasonable expected time commitment of ‘15 – 20 minutes total’, stated in the invitation and instructions for the experiment, is designed to encourage participation. The total number of stimuli used is therefore determined by the number of replications and the average time taken.

3.1.7 Replications

Adhering to guide values and recommendations for the number of n observations guards against the probability of a Type II error, β, 'the chance of missing a true difference'.[139(p104)] An increase in n without extra recruiting can be achieved by presenting the subject with the same trial more than once, where the power of the test increases considerably with a small number of replications.[140] A common test statistic for repeated difference tests is the sum of all correct assessments, the validity of which is the matter of some disagreement in the literature.[141–144] Kunert asserts that 'if the null hypothesis is true and the experiment is properly randomised and properly carried out, then all assessments are independent and all have the same success probability…'.[145(p385)] Therefore, items are randomised as standard within each trial, and all trials for each object within a session are presented in randomised order.

Four replications are used with regard to the maximum session duration and the required number of stimuli,[146] where Bech recommends between two and four replications as 'usually necessary in order to achieve the desired statistical reliability'.[147(p103)] For an estimated number of participants $\geq$ 31, based upon the prior experiment, four replications yields a fourfold increase in n, aligning with the quote, 'while there is no correct sample size, recommendations are for a size ranging between 100 to 200 … but with an absolute minimum of 50 respondents'.[148(p444)] In accordance with the ISO 10399:2018 for duo-trio tests and ISO 5495:2016 for paired comparison tests each replication is determined as independent.[142,149,150]

For the audible polarity experiment the duo-trio test replications are configured as a *balanced-referenced*.[137] In the case of four replications, item A is assigned to be the *Reference* for two of the triads, item B is assigned the *Reference* for the remaining two triads, and each replication is presented in random order within the randomised trials. To avoid confusion the experiment is separated into two sections, one section for the duo-trio tests, and another for the paired comparison tests. The two sections are presented in randomised order with a repeated instruction page at the beginning of each section.

A planned small-scale pilot study was made, one of the purposes of which was to determine if the number of replications was suitable, in terms of statistical analysis and average total length of time per session.[107]

3.1.8 Semantic Perception

A search of the authors' PDF[151,152] research database for the phrases 'listening experiment' or 'listen experiment' returns 81 results, whilst the phrases 'listening test' or 'listen test' returns 396 results.[153] Figure 7.4 in Appendix Section 7.2, shows how the phrases 'listening test' and 'listening experiment' have occurred in the *English 2019* corpus of books from 1900 to 2019.[154]

Feedback and discussions with colleagues and associates from the associated prior listening "test" led to a possible conclusion that the phrase "listening test" may imply an examination of a subjects "listening ability".[7] The word "test" may invoke what Kahneman describes as effortless "System 1" *intuitive* thinking, rather than "System 2" *deliberate* reasoning,[155] where '...System 1 is more influential than your experience tells you, and it is the secret author of many of the choices and judgments you make'.[156(p13)] In terms of Tversky and Kahneman's *Representativeness Bias*[157] the individual may make an immediate judgement, a heuristic principle, of what the word "test" entails.[158]

Therefore, the term "listening experiment" is preferred here. The word "test" was not used to describe the experiment, did not appear in any instructions, or within the experiment itself. Attention and care must be taken, throughout the listening experiment, to minimise the risk of subject bias. For example, if the subject believes their "hearing ability" is being evaluated, and if their profession relies upon it, they may not participate through perceived risk of judgment.

3.1.9 Questionnaire

A questionnaire is added as the final webpage of the online experiment, including a request for an optional email address, and opportunity to leave written feedback. The questions are determined by categories of assessor types[120] and experimental conditions, where subset groups enable the opportunity for possible deeper understanding of results.[101,159,160] Assessor categories are discussed in Section 3.2.2 Assessor Types.

The number of categories is limited with regard to the anticipated number of participants and minimum subset group size (see Section 4.1.1 Cluster Analysis). Headphone and earphone types are brought into four broad categories using common terms: *Closed-back* (circum-aural headphone) ; *Open-back* (circum-aural headphone); *In-Ear* (insert earphone); *Moulds* (custom moulded insert earphone).[81,89] See Figure 7.3 in Appendix Section 7.2 for diagram examples. Details of the questionnaire may be found in Table 7.9 located in Appendix Section 7.1.

3.2 SUBJECT SELECTION

3.2.1 Sample Size

Participants were self-selecting from direct email to the authors' colleagues, friends, and associates, and the University of Derby faculty. Regular announcements were posted on email discussion groups specialising in audio, acoustical society and audio engineering social media pages, and further. The main study was active, as planned, from 2021-07-01 until 2021-07-31. An incentive of six (6) gift vouchers were offered, to be randomly selected from those participants who entered their email address, otherwise participants were not financially compensated for their time.

'The greater the sample size, the greater the power and the more sensitive the experiment in detecting treatment differences in the population'.[161(p70)] Recommendation ITU BS.1284-2 *General methods for the subjective assessment of sound quality* states 'the minimum number of expert listeners should normally be ten, whilst the minimum number of non-expert listeners should normally be twenty'.[162(p3)] The number of self-selected participants completing the experiment could not be determined beforehand.

Table 3.1: Sensory Assessor Types[120,163]

TYPE OF ASSESSOR	DESCRIPTION
"sensory assessors"	Any people taking part in a sensory test. They can be "naïve assessors" who do not meet any particular criterion, or "initiated assessors" who have already participated in sensory tests
"selected assessor"	Chosen for their ability to perform a sensory test
"expert sensory assessor"	In the general sense, a person who, through knowledge or experience, has competence to give an opinion in the fields about which he/she is consulted

3.2.2 Assessor Types

The ITU-R BS.1387-1 standard for objective measurements of perceived audio quality is based upon an 'average expert listener',[164(p74)] and recommends the use of 95% confidence intervals in the statistical analysis. However the aim of this field experiment was not to poll solely the 'expert assessors' and 'experts', as recommended or preferred in various listening test standards.[101,162,164–166]

The perception and preferences of random selected listeners of commercially available stereophonic programme material from all categories of the population is of interest in this study. Table 3.1 shows a summary of assessor categories employed in sensory analysis, defined in ISO 5492:2009 +A1:2017 *Sensory analysis—Vocabulary*,[163(pp4, 6)] as applied to the food industry, and recommended for use in the field of audio.[107(p110),167(p4)] See table in Figure 7.2 located in Appendix Section 7.1 for more detail.

Further categorisation, defined by the Questionnaire, was led by auditory neuroscience research indicating that musical training boosts auditory and cognitive skills,[168–170] where 'The active, repeated engagement with sound demanded by music making augments the neural processing of speech, eventually cascading to listening and language'.[171(p287)]

3.3 STIMULI SELECTION AND PROCESSING

3.3.1 Digital Audio Workstation

The digital audio workstation software *REAPER*[172] was used throughout to edit, process, and audition the programme excerpts and synthetic signals. Audio analysis was performed using the software *ARTA*,[173] *EASERA Pro*,[174] *LIVE-CAPTURE Pro*,[175] and *AFMG SYSTUNE*.[176]

All programme excerpts used were first generation uncompressed Waveform Audio File Format (WAV) or Apple Lossless Audio Codec (ALAC) files sourced from compact disk or digital download.

3.3.2 REAPER Mix Configuration

Matching crosstalk output levels for PAN and MS methods to allow comparison were calibrated according to the relative difference in output level for a signal input to one channel only. See Figure 7.29 located in Appendix Section 7.2 for an example mix window view for a 1 kHz sine source routed to Left only. In this instance a Mid level offset of +6.04 dB is matched by a pan control positions of 59 % using the −6.02 dB pan law.

The stock *REAPER* plugin *JS: Mid/Side Encode* was used to encode and decode.[172] The plugin's digital signal calculations are consistent with Blumleins' original patent, 'two sets of impulses are further transmitted consisting in [sic] half the sum and half the difference respectively of the original impulses'[97(p10)]:

$$Mid_{\text{output}} = \left(Left_{\text{input}} + Right_{\text{input}}\right) \times 0.5$$

$$Side_{\text{output}} = \left(Left_{\text{input}} - Right_{\text{input}}\right) \times 0.5$$

Dirac impulses were used to verify REAPER's sample accurate synchronisation between the MS and PAN mix renderings.

3.3.3 Synthetic Signals and Programme Excerpts: Audible Polarity Experiment

Table 3.2: Synthetic Signal and Programme Excerpts for Polarity Experiment

ITEM	TITLE	ARTIST	ALBUM	YEAR
POL3	300 Hz Gauss envelope sine burst, 4 sine periods	n/a	n/a	
POL4	Billet Doux: The Complete Unedited Version	unknown	BBC Radio Group 2 Acoustics Test CD	unknown
POL8 [not in pilot study]	Excerpt From MARINA (Normal Polarity)	The David Chesky Band	Various – Chesky Records Jazz Sampler & Audiophile Test Compact Disc Vol. 1 / Chesky Records / CHESKY JD37 / US	1990
POL7 [training section]	Blues In The Night	Frank Sinatra	Frank Sinatra Sings For Only The Lonely / Mobile Fidelity Sound Lab / MFSL 1-326 / US / 2009	1958

Table 3.2 lists the programme excerpts and synthetic signal used to create the respective Reference and Inverted Polarity items for comparison. Seventeen (17) programme excerpts of commercially available stereophonic material were shortlisted from a wide selection of prospective stimuli. Appendix Section 7.6 contains a list of auditory demonstration albums and test signal sources included in the audition process for this experiment.

With reference to the prior crosstalk experiment, multiple alternative tone burst test signals of varying frequency, number of sine periods, and envelope windowing were generated and trialled by the author, auditioned using a single loudspeaker, stereo loudspeakers, multiple headphone and earphone models and types. The stereo loudspeaker listening arrangement used the optimal base angle of 60° and optimal listening distance of $LR_{dist} \times 0.9$, where LR_{dist} is the horizontal distance between the Left and Right loudspeakers.[23(p10)]

Figure 3.1 and Figure 3.2 show a time record for the Left channel of the synthetic signal POL3 used in the experiment, a 300 Hz Gauss envelope tone burst of 4 sine periods, the Right channel is identical. Linkwitz utilises a similar signal, with Blackman envelope windowing, for audible room reflection and driver distortion investigations (Linkwitz, 1980,2019b).[177,178] The 300 ms spacing between signals is empirically determined to be at the boundary of the reported theoretical psychophysical threshold for a pair of signals to be considered independent.[179(p57)] With this spacing the second signal may reinforce the initial auditory perception. The pair of spaced tone bursts are repeated after 685 ms. Tone bursts were generated using the Signal Generation and Recording function in *ARTA*.[180(p90)]

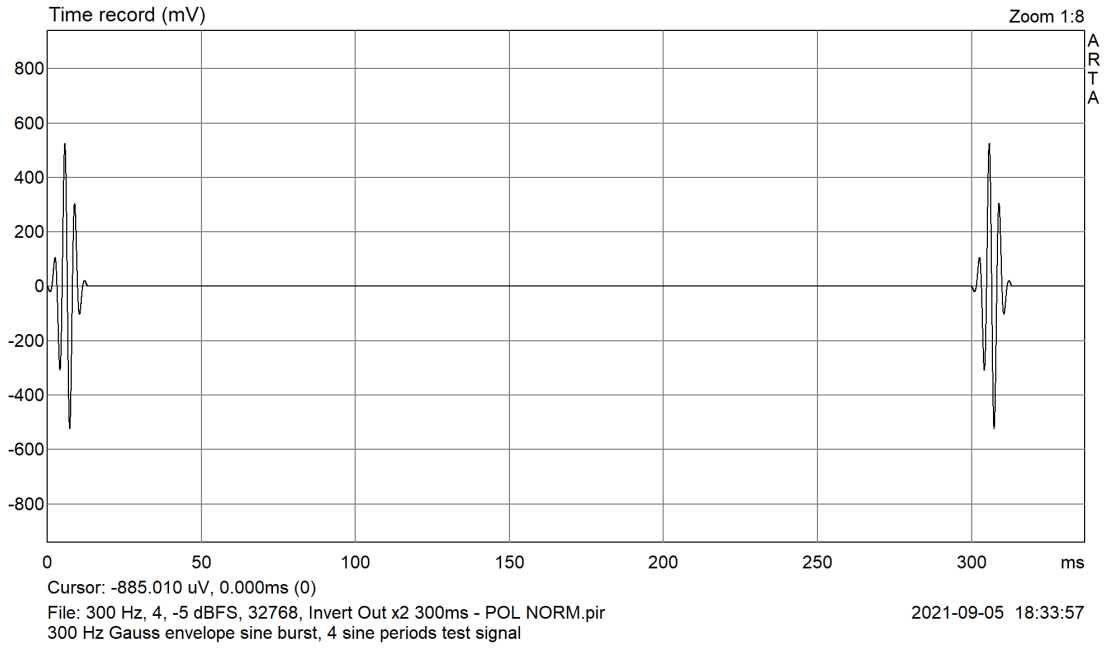

Figure 3.1: *POL3* test signal, a 300 Hz Gauss envelope tone burst, 4 sine periods. Zoom 1:8

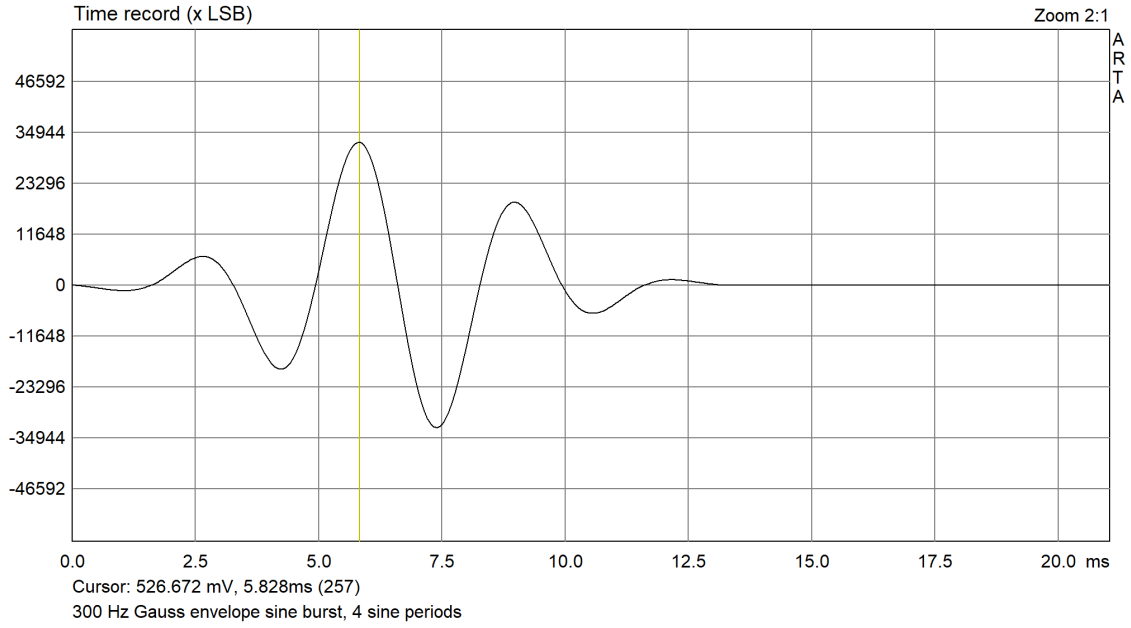

Figure 3.2: *POL3* test signal, a 300 Hz Gauss envelope tone burst, 4 sine periods. Zoom 2:1

3.3.4 Programme Excerpts: Crosstalk Experiment

Table 3.3: Programme Excerpts for Crosstalk Experiment

ITEM	TITLE	ARTIST / GROUP	ALBUM	YEAR
MP6	It's Too Late	Carole King	Tapestry / Ode Records, Epic, Legacy / EK 65850 / US / 1999	1971
MP12	What'd I Say (Parts 1 And 2)	Ray Charles	The Definitive Ray Charles / Warner Strategic Marketing International, Atlantic, Rhino Records / 8122 73556-2 / Europe / 2001	1959
MP18	I Can't Go For That (No Can Do)	Daryl Hall & John Oates	Ultimate Daryl Hall + John Oates / BMG Heritage / 82876573552 / US / 2004	1981
MP22	Rocket Man (I Think It's Going To Be A Long, Long Time)	Elton John	Honky Château / Mercury / 528 162-2 / UK & Europe / 1995	1972
MP47	Stayin' Alive	Bee Gees	Saturday Night Fever (The Original Movie Sound Track) / Mobile Fidelity Sound Lab, Polydor / UDCD 716 / Russia / 1998	1977
MP15 [training section]	Prelude No. 2 In C Minor From "The Well-Tempered Clavier", Book I	Glenn Gould	Thirty Two Short Films About Glenn Gould (Original Motion Picture Soundtrack) / Sony Classical / SK 46686 / US / 1994	1963
MP23 [pilot study only]	You Shook Me All Night Long	AC / DC	Back In Black / EMI / 495 1532 / UK & Europe / 1998	1980
MP24 [pilot study only]	Time	Pink Floyd	The Dark Side Of The Moon / EMI / CDEMD 1064 / UK / 1994	1973

Table 3.3 lists the programme excerpts used to create the respective MS and PAN paired items for comparison. One-hundred-and-one (101) programme excerpts of commercially available stereophonic material were auditioned by the author, using stereo loudspeaker listening arrangement previously described, and multiple headphone and earphone models and types. Prospective excerpts were prioritised according to highest individual units and album units sold, by territory, year, decade and all-time.[181–184]

The average between fifteen to twenty-minute target to complete the online experiment limits the number of trials, replications, and number of stimuli. A single crosstalk level was empirically determined, based upon the research undertaken for the prior experiment and further observations.[185,186] Figure 3.3 shows measured Interaural Level Differences (ILD) for a 'typical listener seated in a 60° stereo triangle in the professional listening room'.[187(p9)] Extensive trials with programme material and synthetic signals led to a *Mid* $(L + R)$ level of +6 dB, equal to a crosstalk level difference, for a signal input to one channel only, of 9.5 dB.

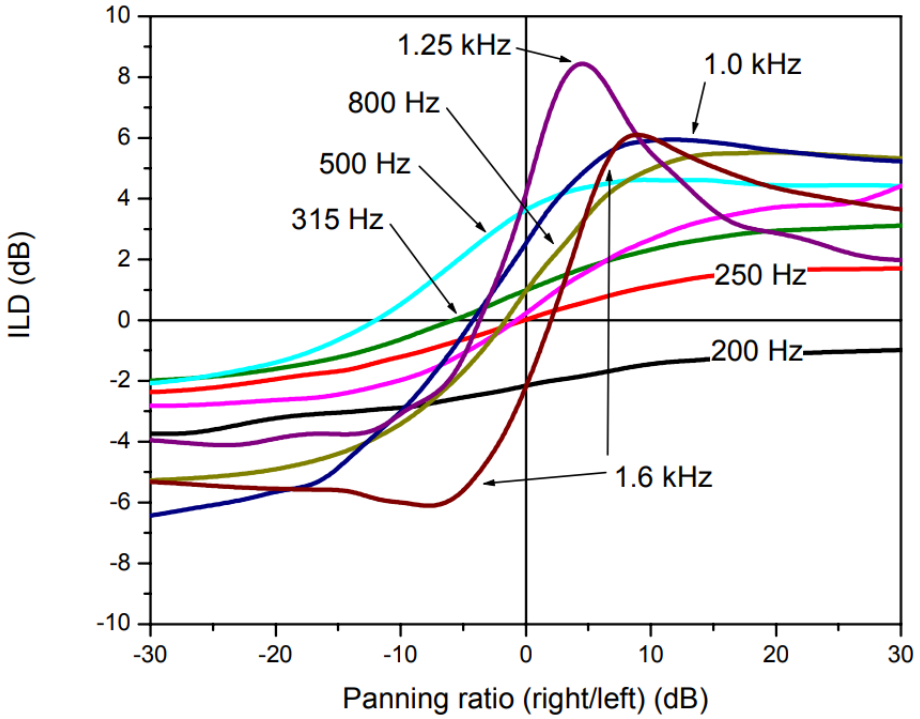

Figure 3.3: 'ILDs for a [typical] listener and 60° stereo triangle'.[187(fig19)] Reproduced from Benjamin E, 2006.

3.4 CONFIGURATION AND CALIBRATION OF THE EXPERIMENTAL SET-UP

3.4.1 Listening Test Software Platforms

The experiment platform criteria:

- Accessible through a web browser, with no additional software required
- Suitable viewing and functionality on a personal computer, tablet, and smartphone
- Proven stable operation and data result recording

- Configurable test methods, i.e., duo-trio and paired comparison

Various listening experiment platforms were investigated.[188–195]

A wiki of remote testing platforms is compiled and maintained by the Acoustical Society of America, initiated in Spring 2020.[196]

3.4.2 *webMUSHRA* Configuration

The server-based web application *webMUSHRA* was used for the experiment, accessible through a web browser, and able to meet all required criteria of the experiment.[102] Minor modifications were implemented for desired presentation views across all screen sizes and browser types, and for custom test functionality.

webMUSHRA utilises the *Web Audio API*[197] a high-level Application Programming Interface for processing and synthesizing audio in web applications offering client side processing, with performance defined by the client device and consistent across different web connection conditions. *webMUSHRA* sets a maximum item duration to 12 seconds, in accordance with ITU-R BS1534-3,[101] calculated to avoid fatigue, increase stability and robustness of listener responses, and limit total session duration. The test interface features looped playback with synchronised switching between the items. A trial must be completed before proceeding forward in the experiment.

Example web browser views from the main study are presented in Figure 7.23 to Figure 7.24, located in Appendix Section 7.2.

3.4.3 Loudness Normalisation

The loudness levels[198] for samples were consistent with EBU R 128,[199] which incorporates *Recommendation ITU-R BS.1770-4*.[200] Programme excerpts for Short-form Content, a programme typically shorter than 30 seconds but up to approximately 2 minutes, use the approach recommended by *Supplement 1 to EBU R 128*,[201] due to an insufficient number of Short-term (3 sec) Loudness value data points. The ITU 1770 / EBU R 128 meter in *RME Digicheck* software[202] was used to normalise all items to a 'Programme Loudness' of -23.0 LUFS,[203] 'Maximum Short-term Loudness' of -18.0 LUFS and 'Maximum True Peak Level' of -1 dBTP.[204]

3.4.4 Level Calibration Stimulus

A level calibration stimulus is used at the start of the experiment to allow subjects to adjust appropriate levels for their playback equipment.[205] This ensures subject awareness of being in control of the playback level. The principal aim being for all subjects to select a similar playback level with their personal equipment.

The excerpt used is an edit of 'Loudness Calibration pattern for playback volume' from the compact disc *Demonstrations to accompany Bregman's: Auditory Scene Analysis - The perceptual organization of sound*.[206] The signal consists of a 1.5 kHz sine tone at −40 dBTP for 900 milliseconds, followed by 100 milliseconds of pink noise low passed at 8 kHz, at −17 dBFS RMS, −12 dBTP. This sequence is looped twenty times. The complete sequence gives a maximum 'momentary' ('M') scale level of 0.0 LU (-23.0 LUFS).

3.4.5 Left-Right Balance Stimulus

It has proven sensible to offer a Left-Right balance stimulus for subjects to verify correct operation of the playback equipment (Digby, 2021; mcdermottLab, 2021).[7,207] The balance stimulus consists of a vocal description, followed by a tone burst signal sequence to the Left channel only, then repeated for Right channel only. The Linkwitz tone burst sequences consist of 4-cycle cosine Blackman envelope tone bursts of 100 ms duration descending fast from 12800 Hz to 100 Hz in twenty-two steps, followed by 500 ms tone bursts in a slow ascent from 100 Hz to 12800 Hz in the same twenty-two steps.[208] The sequence ends in a 100 Hz tone burst followed by another at 1000 Hz. All tone bursts for this sequence calibrated to −9.5 dBTP using RME Digicheck.

The sequence may highlight the non-stationary localization of a dichotic signal with relation to frequency. See Blauert for further discussion, and Figure 7.22 located in Appendix Section 7.2.[63(pchap.2.1)]

3.4.6 Training Stage

An example of a duo-trio test for audible polarity and paired comparison for the crosstalk experiment are presented before the experiment trials to familiarise subjects.[107(p314)] A key aspect of the remote field study, being unsupervised, is simplicity of the test method. The nature of the experiments does not allow for easily discernible programme excerpts to be used for the training stage. It is considered important for the training stage to accurately represent the same difficulty level as the experimental trials that follow. The results of the training stage trials are not included in any statistical analysis.

3.5 PILOT AND MAIN STUDY

Six respondents from a pool of expert assessors completed a pilot study. Where an individual subject's choices for replications within a trial were equal to chance, 1/ 2, the programme excerpt was discarded from the main study (*MP23* and *MP24*). Average session duration required reducing the total number of trials for the main study. A higher-than-expected discrimination agreement between subjects for the audible polarity trials led to the addition of a new audible polarity programme excerpt (*POL8*) to the main study.

Follow-up enquiries included adjustments made in playback level, by one subject, after having completed the Level Calibration at the start of the experiment. Further investigation led to re-normalising items using the unmodified R 128 loudness method, with use of the *'Loudness Range'* measure achieved by looping programme excerpts for $>$ 3 minutes.[203(p43)] Changes of less than $\pm$1 dB resulted. Minor modifications were made to *webMUSHRA* pages according to feedback suggestions.

The main study ran from July 1st to July 31st, 2021. The author responded to feedback comments for the duration, where appropriate, to express thanks for participating and to arrange to reveal the nature of the experiments after report submission where desired. It is considered good practice to minimise expectation bias by not revealing details of a listening experiment, outside of necessary instruction.[107(p96)]

4 RESULTS

The approach, methods, and algorithms of the statistical analysis were determined by recommended practices prior to the data analysis.[145,209–216,216–218] All observations were retained,[148(p49)] and initial statistical methods adhered to throughout.[219–221]

4.1 STATISTICAL ANALYSIS

The free software environment *R*[222] was used for the statistical computing and graphics output of the *webMUSHRA* comma-separated values (CSV) data record.[223] Eighty-six (86) male and ten (10) female participants completed the experiment; a sample size of $N = 96$. Table 4.1 shows the participants' age distribution. Figure 7.5 located in Appendix Section 7.2 offers a distribution plot.

Table 4.1: Summary of Sample Ages (years)

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
9.00	33.75	47.50	44.81	55.00	73.00

4.1.1 Cluster Analysis

Subset groups from the sample were investigated for the statistical analysis, listed in Table 4.2.[148(p267)] Intersecting set graphs generated using the UpSet method from Lex et al.,[224] using the *R* package `UpSetR`,[225] allowed rapid correlations to be visualised. See Figure 7.6 and Figure 7.6 located in Appendix Section 7.2.

Appropriate groupings were determined by size recommendations, 'values of $n < 24$ are usually not recommended for a duo-trio test for a difference' ISO 10399,[137(p7)] and 'values of $n < 18$ are usually not recommended for paired difference test' ISO 5495.[132(p8)] Both ISO standards contain tables for matching desired α and β values to a sample size, or the reverse. See Figure 7.30 in Appendix Section 7.4 for the adapted Schlich[226] table for determining the number of assessors from ISO 10399:2018 *Sensory analysis – Methodology – Duo- trio test*.[137(p12)]

Table 4.2: Sample Groups for Cluster Analysis

GROUP	N	n	**DESCRIPTION**
All	96	384	All participants
age <48	48	192	Less than 48 years of age
age ≥ 48	48	192	48 years of age and higher
intersect	24	96	Answered 'yes' to all of the following: 'musical training / play and instrument / sing' + 'can speak more than one language' + 'have you taken part in a listening experiment before' + 'background in professional audio'
!intersect	72	288	Not in *intersect* group
closed 'phone	55	220	Used closed-back headphones for experiment
!closed	41	164	Did not use closed-back headphones for experiment
multilingual	48	192	'Can speak more than one language' = Yes
listen test exp	56	224	'Have you taken part in a listening experiment before' = Yes
musical train	64	256	'Do you have any formal musical training / play and instrument / sing?' = Yes

4.1.2 Wilcox *R* Package

Wilcox[227] supplies the custom *R* function `binom.conf` used for the analysis, applying the Schilling-Doi method for $n < 35$,[228] and an Agresti-Coull method for $n \geq 35$.[229] Wilcox states the essence of the Agresti-Coull method as an approximation of the score method derived by Wilson (1927),[230] supported by Zou et al. (2009).[231] Examples of the *R* code and explanation of the Agresti-Coull method are located in Appendix Section 7.5.3.

4.1.3 Results Summary – Polarity Experiment Duo-trio Test

Table 4.3 shows a summary of notable results, with respect to the Type I error probability α (rejecting the null hypothesis when it is true). For binomial confidence intervals that do not include 0.500 this allows the rejection of the null hypothesis, $H_0 : P_{d_t} = 0.5$, where P_{d_t} is the long-run probability of the population making a correct selection when there is no perceptible difference between the items, one in two.[214(p122)] In the first row of Table 4.3 the *POL3* trials for the "all" group show with 99.9% confidence that values for this sample group will be within 0.502 and 0.665, inferring a detectable difference between the *POL3* reference item and its inverted polarity form. See Table 7.1, Table 7.2, and Table 7.3 located in Appendix Section 7.1 for full results.

Table 4.3: Polarity Experiment Duo-Trio Test: Results Summary

TRIAL	GROUP	n	99.9% CI	99% CI	95% CI	90% CI
POL3	all	384	(0.502, 0.665)	(0.520, 0.649)	(0.536, 0.634)	
POL4	all	384		(0.502, 0.631)	(0.518, 0.616)	
POL8	all	384			(0.497, 0.596)	(0.505, 0.588)
POL8	!intersect	288		(0.504, 0.652)	(0.522, 0.635)	

4.1.4 Results Summary – Crosstalk Experiment Paired Comparison Test

Table 4.4 shows a summary of notable results, with respect to the Type I error risk probability α (alpha). In the crosstalk paired comparison trials a preference for the PAN method registered as a score of zero (0), preference for the MS method a score one (1). The first row of Table 4.3, the *MP6* trials for the "!intersect" group, show that with 90% confidence the values for this sample group will be within 0.400 and 0.496, inferring a preference for the item processed with the PAN method. The binomial confidence intervals do not include 0.500 allow the rejection of the null hypothesis $H_0 : p = 0.5$,[232(p144)] where the proportion of responses for one of the samples, p, is equal to 1/2.[107,132(p3),227] This would correspond to the equal probability of choosing either of the two items.

In the last row the *MP47* trials for the "all" group show with 90% confidence that values for this sample group will be within 0.502 and 0.586, inferring a preference for the item processed with the MS method. See Table 7.4, Table 7.5, Table 7.6, Table 7.7, and Table 7.8 located in Appendix Section 7.1 for full results.

Table 4.4: Crosstalk Experiment Paired Comparison Test: Results Summary

TRIALS	GROUP	N	n	95% CI	90% CI
MP6	!intersect	72	288	(0.392, 0.506)	(0.400, 0.496)
MP12	intersect	24	96	(0.313, 0.506)	(0.327, 0.490)
MP18	All	96	384	(0.445, 0.545)	
MP22	All	96	384	(0.458, 0.557)	
MP47	All	96	384	(0.494, 0.593)	(0.502, 0.586)

4.1.5 Voucher Incentive: Winner's Draw

After the closing date of the experiment six participants were selected from the dataset utilising a Mersenne Twister pseudorandom number generator.[233] The *R* code, using its own initialization method, is reproduced in Appendix Section 7.5.2.[222]

4.1.6 Volume Setting Adjustment

Approximately 65 % of participants did not adjust the volume level of their playback device after the calibration stage, compared with approximately 60 % of thirty-one participants in the prior experiment. A different stimulus was used for each.

5 DISCUSSION

5.1 DISCUSSION — AUDIBLE POLARITY

5.1.1 Inferred Audibility of Polarity

The statistical results from the *POL3* trials, the synthetic tone burst stimulus, infers with 99.9% probability that the sample population correctly discerns the difference between the stimulus and its polarity inverted form under these conditions. This probability level indicates very strong evidence that a difference was apparent.[137(p11)]

From the *POL4* trials, the male spoken word excerpt, there is strong evidence that a difference was apparent for the sample population under these conditions. From the *POL8* musical ensemble excerpt trials there is slight evidence of audible polarity discernment for the sample population under these conditions, together with strong evidence of discernment for the subset group *!intersect* (see Table 7.3 in Appendix Section 7.1).

The results are of interest, with consideration of the field experiment conditions and the literature surrounding the phenomenon.

5.1.2 Follow-up Study — Audible Polarity

> For large sample sizes, we only need a small proportion of discriminators to push us over the statistically critical proportion for significance … although we may have found a significant difference, not everyone can discern it.
>
> — Harry T Lawless and Hildegarde Heymann, 2010.[139(p107)]

The experimental design balanced the number of stimuli and replications for both experiments with the maximum suggested length of time, whilst attempting to maintain the subjects' engagement with the exercise. A different approach may be to use the duo-trio test method but with $\geq$ 10 replications instead of four. This may allow a useful estimation of the inferred proportion of discriminators in the population.[211]
See Figure 7.1 located in Appendix Section 7.1.

Schlich, Dacremont and Brockhoff[234] emphasize the value of using replicates over earlier findings[142] by showing that a single subject having done ten replicates may bring to an experiment the same amount of useful information as five subjects having done the experiment once. The risk, in this case, would be for a modest number of these 'single subjects' to not be in the group of discriminators.

Historically a common minimum number of participants is $n = 40$, based upon a normal sampling distribution, however Wilcox notes one of the circumstances where the $n = 40$ rule breaks is with the binomial distribution where the probability of success, p, is close to '0' (zero) or '1' (one).[216(p122)]
The minimum number of participants is therefore recommended to be $\geq$ 100, in accordance with arguments made earlier in this report, and where the greater the sample size the closer the approximation is to the sample mean.[235]

5.1.3 Polarity Standards

In *The New Stereo Soundbook* from 1998, Streicher and Everest state that 'Unfortunately, no standards yet exist for maintaining absolute polarity throughout the entire signal chain'.[73(p7.5)] The following selection of standards from four organisations recommend preserving absolute polarity through the reproduction chain:

1. Audio Engineering Society (2017) *AES26-2001 (stabilized 2017) AES recommended practice for Professional audio interconnections — Conservation of the polarity of audio signals.* AES.
2. European Broadcasting Union (1998) *R50-1998 Conservation of the polarity of audio signals in radio and television production installations.* EBU.

3. International Electrotechnical Commission (1994) *IEC 1213:1993 Analogue audio recording on video tape — Polarity of magnetization*. CENELEC.

4. Society of Motion Picture and Television Engineers (1994) *RP 134:1994 - SMPTE recommended practice - polarity for analog audio magnetic recording and reproduction*. SMPTE.

IEC 60268-7 *Part 7: Headphones and earphones*[81(p15)] requires polarity to adhere to the following from IEC 60268-2: 'A terminal of an electroacoustic transducer has positive polarity when ... a positive instantaneous voltage at that terminal produces an outward movement of a diaphragm'.[16(p8)]

The author performed field measurements estimating the phase response of a selection of headphones and earphones, inconsistencies were found. For centred graphs of the unwrapped phase response, with a maximum y-axis range of 360°, the mid-point is seen to be 0° for seven of eleven devices, with ambiguity within. The graphs are presented in Figure 7.8 to Figure 7.18 located in Appendix Section 7.2.

5.1.4 Audio Waveform Asymmetry

The audibility of a difference between a signal and its inverted polarity form is of interest with regard to the asymmetry of naturally occurring waveforms, such as speech. With reference to the literature regarding auditory pattern recognition (reviewed in Section 2.1.4: Neuroscience and Auditory Perception), perceived change in pitch,[12(p246)] and Kraus and Nicol's note that the Frequency Following response 'might be considered a series of repeated transients'.[40(p177)] A naturally occurring waveform and its equivalent inverted polarity form may not be considered interchangeable in auditory perception.

Empirically observed spatial effects with inverted polarity may be associated, where a signal reproduced with a loudspeaker in absolute polarity may appear to 'envelope' or 'surround' the listener and inverting the polarity may result in an image shift, with the perceived source origin moving to approximately behind the loudspeaker cone.

5.2 DISCUSSION — CROSSTALK

5.2.1 The “missing 6 dB”

Of interest is the correlation between the empirically derived addition of 6 dB to the *Mid* channel for the crosstalk experiment and the discussion across a number of papers of ‘the “missing 6 dB” paradox’,[236(p650)] as quoted in *The case of the missing 6 dB*.[237–243] It is stated as the necessity for ‘approximately 6 dB more sound pressure level on the subject’s eardrum’[236(p650)] for earphone loudness judgements of low frequencies in comparison with a loudspeaker in free field as the source, originally reported by Beranek (1949).[244]

Without explicit details of the test environments and conditions across the different experiments some discrepancies may not be accounted for.[2(p654)] The dominance of the free-field or diffuse-field calibration for headphones as the considered equivalent to the loudspeakers’ room response may also contribute.[245(p3)] Importantly, Denk *et al.* (2021), with reference to Rudmose (1982),[236] Zahorik and Wightman (2001),[246] and Sivonen and Ellermeier (2006),[247] report a large effect of recorded loudness mismatch from ‘small acoustic and non-acoustic differences in an anechoic reference environment’.[243(p1)] Their conclusion being that ‘the binaural listening mode and the interaural coherence – which is usually also connected to the apparent source width … appears to play an important role in the differential judgment of loudspeaker vs. headphone presentation’.[243(p2)]

5.2.2 Difference of the Difference

A comparison of the MS and PAN crosstalk method transfer functions were investigated by first processing the difference between Left and Right outputs of each method, respectively, then comparing the difference between the two differences. The results show the MS and PAN crosstalk methods to not be identical. See Table 7.10 located in Appendix Section 7.1.

This is not unexpected with consideration of the inferred discernible difference results for *MP6, MP12* and *MP47* found here, and the different ratings received for both crosstalk methods in the prior experiment. However, further investigation of the two methods is of interest, including the commutativity of the mathematical processes and spectrogram analysis.[59(p184),248]

5.3 DISCUSSION — METHOD

5.3.1 Allowable Guesses

The use of tied judgements, the third option in a pairwise comparison such as a “No difference” response, is reported by Perez-Ortiz and Mantiuk as a ‘controversial issue … still disputed and researched’,[249(p17)] and further considered an ‘open research question’.[250(p3174)]

For this experiment instructions presented to participants in the invitation and web application introduction did not give specific guidelines for guesses, outside of stating ‘it is not a test … there are no wrong answers’ (see Appendix Section 7.3). A cause for concern was a minority of comments in the *Feedback* section of the questionnaire which indicated subjects, if unable to discern a difference, may have questioned their hearing acuity. This was immediately addressed via direct communication in each case.

The duo-trio test is a forced-choice procedure with a “no difference” or “I don’t know” option forbidden, however, an individuals’ disposition may not allow them to continue the experiment if feeling unsure or deceptive in making a guess. In similar fashion the paired comparison test, used for the crosstalk experiment, is a forced choice test between two alternatives, sometimes referred to as a directional difference test or a 2-AFC test, *Alternative Forced Choice*.[132(p1)] It is possible that a participant may doubt their own discernment yet successfully identify their chosen item variant for repeated trials. A record of incomplete sessions is unknown as results are only recorded upon completion of the questionnaire.

5.3.2 Time Management

At the outset, the allocated time for the entire research project was organised according to the following, in *Perceptual Audio Evaluation – Theory, Method and Application*: ‘about one-third of the time will be spent preparing the experiment, one-third for administrating the pilot and main experiment and the remaining one-third for analysis and reporting’.[107(p303)] This was proven to be the case.

The Gannt chart,[251] a plan of works prepared at the outset, and a weekly progress record is available at the link: 100559332 7EJ998 IES gantt 20210526

6 CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

A web-based field listening experiment was completed by ninety-six subjects. The audibility of acoustic polarity with biphonic reproduction was investigated for three stimuli. The second section of the experiment investigated subject preference and discernible difference with biphonic reproduction for five stimuli using two crosstalk methods, PAN (amplitude panning) and MS (Mid/Side).

1. There is very strong evidence to indicate, for the study population, that a difference may be apparent between the synthetic stimulus *POL3* and its polarity inverted form using biphonic reproduction.
2. There is strong evidence to indicate, for the study population, that a difference may be apparent between the male spoken-word stimulus *POL4* and its polarity inverted form using biphonic reproduction.
3. There is moderate evidence to indicate, for the study population, that a difference may be apparent between the music excerpt *POL8* and its polarity inverted form using biphonic reproduction.
4. There is slight evidence to indicate, for the study population, a preference for the MS crosstalk method with programme excerpt *MP47* using biphonic reproduction.
5. There is slight evidence to indicate, for a subset group of the study population, a preference for the PAN crosstalk method with programme excerpts *MP6* and *MP12* using biphonic reproduction.
6. There is evidence to indicate, for the study population, that no preference exists between the PAN and MS crosstalk methods for programme excerpts *MP18* and *MP22* using biphonic reproduction.
7. There is evidence to indicate, for the study population, that a difference may be discernible between the PAN and MS crosstalk methods using biphonic reproduction.
8. The result of the crosstalk experiment, with reference to prior work,[7] allow it to be inferred that the study population's may have a negligible preference between the MS and PAN crosstalk re-introduction methods using biphonic reproduction. Therefore, either method is suitable for commercially available stereo programme material, in this context, and may be preferred to the original Left Right presentation.

6.2 RECOMMENDATIONS

6.2.1 Follow-up Study — Audible Polarity

A follow-up study to replicate the results of the audible polarity experiment is recommended again using the duo-trio test method but with ≥ 10 replications instead of four, and a total of three stimuli instead of eight. This may allow the opportunity for reproducibility of the current results,[252] and a useful estimation of the inferred proportion of discriminators in the population.[211] However, this number of replications may have a detrimental effect upon subject engagement and fatigue.

6.2.2 Follow-up Study — Crosstalk

Further analysis of the output differences of the MS and PAN crosstalk methods are warranted. Spectrogram graphic analysis of stimuli processed with both methods may give valuable insight (see Table 7.10: Difference of the Difference in Appendix Section 7.1). Examination of the magnitude differences alone, at ≤ -57 dB Peak, may not give an accurate representation of the complex waveform interactions that may occur with the crosstalk processes.

6.2.3 Remote Listening Experiment

It has been demonstrated that a correctly designed web-based listening experiment, using paired comparison or duo-trio test methods, may provide useful data from a sample group that does not solely consist of expert assessors within the vicinity of the experimenter's laboratory. Online listening experiments may be considered a viable and useful exercise to investigate the diverse and varied wider population.[102,121,253]

6.2.4 Duo-trio and Paired Comparison Guesses

Future remote field experiments by this experimenter, using the duo-trio or paired comparison test method, will instruct assessors who detect no difference between items to randomly select one of the items and 'indicate that the selection was only a guess in the comments section…'.[137(p5)]

6.2.5 Harman Target Curves and Re-Introduced Crosstalk

The Harman Target curves attempt to address the issue of missing loudspeaker crosstalk paths when using biphonic reproduction by applying an appropriate frequency response (see Figure 7.19 located in Appendix Section 7.2). From the preference results of the current and prior crosstalk experiments, it is suggested that application of either the MS or PAN crosstalk methods applied with a Harman Target curve headphone may be a welcome addition to modern hardware and software music players for commercially available stereo programme material.

6.2.6 Unique Perception

No two people hear the same. By the same token, with a person's hearing adjusted to their emotional and physical states, referenced to the individual's constantly evolving history, it may be said that an individual does not perceive an audio stimulus in the same way twice.

References

1. Hilliard J. Notes on how phase and delay distortions affect the quality of speech, music and sound effects. IEEE Transactions on Audio. 1964;AU-12(2):23–25. https://doi.org/10.1109/tau.1964.1161743

2. Borwick J, editor. Loudspeaker and Headphone Handbook. 3rd ed. Focal Press; 2001.

3. Damaske P. Acoustics and Hearing. Springer; 2008.

4. Toole FE. Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms. Third edition. Routledge; 2018.

5. Stewart I. Expert Mastering Engineers Sound Off on Using Headphones for Audio Mastering. iZotope. 2020 Jan 16 [accessed 2026 Oct 19]. https://www.izotope.com/en/learn/expert-mastering-engineers-on-headphones.html

6. Roginska A, Geluso P, editors. Immersive sound: the art and science of binaural and multi-channel audio. Routledge, Taylor & Francis Group; 2018. (Audio Engineering Society presents).

7. Digby JJ. Headphone Image Distortion and Preserving the Stereo Art: Part One. University of Derby; 2021. 2020-7EJ511: Advanced Surround Sound Systems. https://drive.google.com/file/d/1N_QF47Vm_gJqlukPy7xGVoz0Cs-_5_k9/view?usp=sharing

8. Lorho G. Evaluation of spatial enhancement systems for stereo headphone reproduction by preference and attribute rating. In: Audio engineering society convention 118. 2005. http://www.aes.org/e-lib/browse.cfm?elib=13230

9. Martens WL, Marui A. Spatial character and quality assessment of selected stereophonic image enhancements for headphone playback of popular music. In: Audio engineering society convention 120. 2006. http://www.aes.org/e-lib/browse.cfm?elib=13622

10. Craig JH, Jeffress LA. Effect of Phase on the Quality of a Two-Component, Unsymmetrical, Complex Tone. The Journal of the Acoustical Society of America. 1959 [accessed 2026 Oct 19];31(11):1584–1584. http://asa.scitation.org/doi/10.1121/1.1930342. https://doi.org/10.1121/1.1930342

11. Stodolsky D. The standardization of monaural phase. IEEE Transactions on Audio and Electroacoustics. 1970;18(3):288–299. https://doi.org/10.1109/tau.1970.1162110

12. Greiner RA, Melton DE. Observations on the Audibility of Acoustic Polarity. J. Audio Eng. Soc. 1994 [accessed 2026 Oct 19];42(4):245–253. http://www.aes.org/e-lib/browse.cfm?elib=6950

13. Davis D, Patronis E, Brown P. Sound System Engineering. 4th Edition. Ballou G, editor. Focal Press; 2013.

14. Davis D, Davis C, editors. SYN-AUD-CON newsletter. April, 1979. Vol 6 Synergetic Audio Concepts; 1979. https://www.prosoundtraining.com/wp-content/uploads/2019/04/NLV6N3.pdf

15. Davis D, Davis C. Sound System Engineering. 1st ed. H. W. Sams; 1975.

16. International Electrotechnical Commission. IEC 60268-2:1987 Sound system equipment. Glossary of general terms and calculation methods. CENELEC; 1993.

17. Audio Engineering Society. AES26-2001 (s2017): AES recommended practice for professional audio -- Conservation of the polarity of audio signals (Revision of AES26-1995). 2017.

18. Johnsen C. Proofs of an Absolute Polarity. In: 91st Convention. AES; 1991. http://www.aes.org/e-lib/browse.cfm?elib=5547

19. Vanderkooy J, Lipshitz SP. Polarity and phase standards for analog tape recorders. In: Audio engineering society convention 69. 1981. http://www.aes.org/e-lib/browse.cfm?elib=11959

20. Seneff S. A joint synchrony/mean-rate model of auditory speech processing. Journal of Phonetics. 1988 [accessed 2026 Oct 19];16(1):55–76. https://linkinghub.elsevier.com/retrieve/pii/S0095447019304668. https://doi.org/10.1016/s0095-4470(19)30466-8

21. International Electrotechnical Commission. IEC 1213:1993 Analogue audio recording on video tape — Polarity of magnetization. CENELEC; 1994.

22. Society of Motion Picture and Television Engineers. ANSI/SMPTE 222M-1994: For Television -- Control and Review Rooms -- Monitor System Electroacoustic Response. 1994 [accessed 2026 Oct 19]. https://pub.smpte.org/pub/st222/st0222-1994_stable2006.pdf. https://doi.org/10.5594/SMPTE.ST222.1994

23. European Broadcasting Union. EBU Tech. 3276: Listening conditions for the assessment of sound programme material: Monophonic and two–channel stereophonic. 1998 [accessed 2026 Oct 19]. https://tech.ebu.ch/docs/tech/tech3276.pdf

24. Turner RS. The Ohm-Seebeck Dispute, Hermann von Helmholtz, and the Origins of Physiological Acoustics. The British Journal for the History of Science. 1977 [accessed 2026 Oct 19];10(1):1–24. https://www.jstor.org/stable/4025578. https://doi.org/10.1017/s0007087400015089

25. Helmholtz HLF. On the Sensations of Tone as a Physiological Basis for the Theory of Music. 3rd edn Ellis AJ, translator. Longmans, Green, and Co.; 1895. https://archive.org/details/onsensationsofto00helmrich

26. Brociner V. Power bandwidth, frequency response, and music spectra. J. Audio Eng. Soc. 1965 [accessed 2026 Oct 19];13(4):319–322. http://www.aes.org/e-lib/browse.cfm?elib=1189

27. Yost WA. Psychoacoustics: A brief historical overview. Acoustics Today. 2015 [accessed 2026 Oct 19];11(3):46–53. https://acousticstoday.org/wp-content/uploads/2015/08/Psychoacoustics-A-Brief-Historical-Overview.pdf

28. Monforte J, Karley B, Greiner RA. Comments on -Observations on the audibility of acoustic polarity- and author's reply. J. Audio Eng. Soc. 1995 [accessed 2026 Oct 19];43(3):147–149. http://www.aes.org/e-lib/browse.cfm?elib=7955

29. Winer E. The Audio Expert: everything you need to know about audio. Second edition. Routledge; 2018.

30. Heyser RC. Geometrical Considerations of Subjective Audio. Journal of the Audio Engineering Society. 1974 [accessed 2026 Oct 19];22(9):674--682. http://www.aes.org/e-lib/browse.cfm?elib=2732

31. Suzuki H, Morita S, Shindo T. On the perception of phase distortion. J. Audio Eng. Soc. 1980 [accessed 2026 Oct 19];28(9):570–574. http://www.aes.org/e-lib/browse.cfm?elib=3967

32. Eargle J. Loudspeaker Handbook. Springer US; 2003. https://doi.org/10.1007/978-1-4757-5678-4

33. Geddes ER, Lee LW. Auditory Perception of Nonlinear Distortion. In: Audio Engineering Society Convention 115. 2003. http://www.aes.org/e-lib/browse.cfm?elib=12465

34. Borg E, Counter SA. The Middle-Ear Muscles. Scientific American. 1989 [accessed 2026 Oct 19];261(2):74–81. https://www.jstor.org/stable/24987364. https://doi.org/10.1038/scientificamerican0889-74

35. Zakrisson J-E. Experimental Studies on the Function of the Stapedius Muscle in Man. University of Umeå; 1974. https://umu.diva-portal.org/smash/get/diva2:1152344/FULLTEXT01.pdf

36. Mukerji S, Windsor AM, Lee DJ. Auditory Brainstem Circuits That Mediate the Middle Ear Muscle Reflex. Trends in Amplification. 2010 [accessed 2026 Oct 31];14(3):170–191. https://doi.org/10.1177/1084713810381771

37. Worden FG, Marsh JT. Frequency-following (microphonic-like) neural responses evoked by sound. Electroencephalography and Clinical Neurophysiology. 1968 [accessed 2026 Oct 19];25(1):42–52. https://linkinghub.elsevier.com/retrieve/pii/0013469468900850. https://doi.org/10.1016/0013-4694(68)90085-0

38. Coffey EBJ et al. Evolving perspectives on the sources of the frequency-following response. Nature Communications. 2019 [accessed 2026 Oct 19];10(1):5036. http://www.nature.com/articles/s41467-019-13003-w. https://doi.org/10.1038/s41467-019-13003-w

39. Hoormann J, Falkenstein M, Hohnsbein J, Blanke L. The human frequency-following response (FFR): Normal variability and relation to the click-evoked brainstem response. Hearing Research. 1992 [accessed 2026 Oct 19];59(2):179–188. https://linkinghub.elsevier.com/retrieve/pii/0378595592901143. https://doi.org/10.1016/0378-5955(92)90114-3

40. Kraus N, Nicol T. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends in Neurosciences. 2005 [accessed 2026 Oct 19];28(4):176–181. https://linkinghub.elsevier.com/retrieve/pii/S0166223605000470. https://doi.org/10.1016/j.tins.2005.02.003

41. Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology. 2007 [accessed 2026 Oct 19];118(12):2544–2590. https://linkinghub.elsevier.com/retrieve/pii/S1388245707001939. https://doi.org/10.1016/j.clinph.2007.04.026

42. Majka M. Submilisecond acoustic pulses: effective pitch and Weber-Fechner law in discrimination of duration times. 2014 [accessed 2026 Oct 19]:6. https://arxiv.org/abs/1404.6464

43. Skoe E et al. Human brainstem plasticity: The interaction of stimulus probability and auditory learning. Neurobiology of Learning and Memory. 2014 [accessed 2026 Oct 19];109:82–93. https://linkinghub.elsevier.com/retrieve/pii/S1074742713002396. https://doi.org/10.1016/j.nlm.2013.11.011

44. Porges SW. Polyvagal Theory: A biobehavioral journey to sociality. Comprehensive Psychoneuroendocrinology. 2021 [accessed 2026 Oct 19];7:100069. https://www.sciencedirect.com/science/article/pii/S2666497621000436. https://doi.org/10.1016/j.cpnec.2021.100069

45. Rumsey F. Spatial audio. Focal Press; 2001. (Music technology series).

46. Bregman AS. Auditory Scene Analysis : The Perceptual Organization of Sound. Massachusetts Institute of Technology; 1990.

47. Bregman AS. Auditory scene analysis. In: Smelser NJ, Baltes PB, editors. International encyclopedia of the social & behavioral sciences. Pergamon; 2001. p 940–942. https://www.sciencedirect.com/science/article/pii/B008043076700663X. https://doi.org/https://doi.org/10.1016/B0-08-043076-7/00663-X

48. Guski R. Psychological Methods for Evaluating Sound Quality and Assessing Acoustic Information. Acta Acustica united with Acustica. 1997 [accessed 2026 Oct 19];83:765–774. https://www.researchgate.net/publication/200045182_Psychological_Methods_for_Evaluating_Sound_Quality_and_Assessing_Acoustic_Information

49. Pickering NC, Baender E. Two ears in three dimensions. J. Audio Eng. Soc. 1953 [accessed 2026 Oct 19];1(3):255–261. http://www.aes.org/e-lib/browse.cfm?elib=94

50. Linkwitz S. The-Magic-in-2-Channel-Sound. Linkwitz Lab. 2015 [accessed 2026 Oct 19]. http://www.linkwitzlab.com/The_Magic/The_Magic.htm

51. Vermeulen R. Stereo-reverberation. J. Audio Eng. Soc. 1958 [accessed 2026 Oct 19];6(2):124–130. http://www.aes.org/e-lib/browse.cfm?elib=234

52. Bregman AS. Progress in Understanding Auditory Scene Analysis. Music Perception. 2015 [accessed 2026 Oct 19];33(1):12–19. https://online.ucpress.edu/mp/article/33/1/12/62665/Progress-in-Understanding-Auditory-Scene-Analysis. https://doi.org/10.1525/mp.2015.33.1.12

53. Bregman AS. Auditory scene analysis: hearing in complex environments. In: Thinking in Sound — The Cognitive Psychology of Human Audition. Oxford University Press; 1993. https://doi.org/10.1093/acprof:oso/9780198522577.001.0001

54. Warren RM. Criterion shift rule and perceptual homeostasis. Psychological Review. 1985;92(4):574. https://doi.org/10.1037/0033-295x.92.4.574

55. Bregman AS. Auditory Scene Analysis and the Role of Phenomenology in Experimental Psychology. Canadian Psychology/Psychologie canadienne. 2005 [accessed 2026 Oct 24];46(1):32–40. https://ccrma.stanford.edu/~cc/deck.js/surroundSnd2018/pdf/bregman2005.pdf. https://doi.org/10.1037/h0085822

56. Boring EG. The Physical Dimensions Of Consciousness. The Century Co.; 1933. https://ia600707.us.archive.org/25/items/physicaldimensio031612mbp/physicaldimensio031612mbp.pdf

57. Mamassian P, Knill DC, Kersten D. The perception of cast shadows. Trends in Cognitive Sciences. 1998 [accessed 2026 Oct 19];2(8):288–295. https://linkinghub.elsevier.com/retrieve/pii/S1364661398012042. https://doi.org/10.1016/s1364-6613(98)01204-2

58. Zickler T, Belhumeur PN, Kriegman DJ. Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction. In: Heyden A, Sparr G, Nielsen M, Johansen P, editors. Computer Vision — ECCV 2002. Vol. 2352. Springer Berlin Heidelberg; 2002. p 869–884. (Goos G, Hartmanis J, van Leeuwen J, editors. Lecture Notes in Computer Science). http://link.springer.com/10.1007/3-540-47977-5_57. https://doi.org/10.1007/3-540-47977-5_57

59. Warren RM. Auditory perception: an analysis and synthesis. Cambridge University Press; 2008.

60. Acoustical Society of America. ANSI/ASA S1.1-2013 R2020 Acoustical Terminology. 2020.

61. Eargle J. The Microphone Book - From Mono to Stereo to Surround, A Guide to Microphone Design and Application. 2nd edn Focal Press; 2004.

62. Bock TM, Keele, Jr. DB(Don). The effects of interaural crosstalk on stereo reproduction and minimizing interaural crosstalk in nearfield monitoring by the use of a physical barrier: Part 1. In: Audio engineering society convention 81. 1986. http://www.aes.org/e-lib/browse.cfm?elib=5010

63. Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization. Revised Edition. 1996. https://doi.org/https://doi.org/10.7551/mitpress/6391.001.0001

64. Bauer BB. Broadening the area of stereophonic perception. J. Audio Eng. Soc. 1960 [accessed 2026 Oct 19];8(2):91–94. http://www.aes.org/e-lib/browse.cfm?elib=525

65. Linkwitz S. Room reflections misunderstood? In: Audio engineering society convention 123. 2007. http://www.aes.org/e-lib/browse.cfm?elib=14220

66. Linkwitz S. The Magic in 2-Channel Sound Reproduction - Why is it so Rarely Heard? International Journal of Architectural Engineering Technology. 2015 [accessed 2026 Oct 19];2(2). https://www.avantipublishers.com/index.php/ijaet/article/view/378. https://doi.org/https://doi.org/10.15377/2409-9821.2015.02.02.2

67. Toole FE. Sound reproduction: loudspeakers and rooms. Elsevier; 2008.

68. Steinberg JC, Snow WB. Auditory Perspective -Physical Factors. Transactions of the American Institute of Electrical Engineers. 1934;53(1):12–17. https://doi.org/10.1109/t-aiee.1934.5056476

69. Bauer BB. Stereophonic Earphones and Binaural Loudspeakers. J. Audio Eng. Soc. 1961 [accessed 2026 Oct 19];9(2):148–151. http://www.aes.org/e-lib/browse.cfm?elib=471

70. Wiener FM. On the Diffraction of a Progressive Sound Wave by the Human Head. The Journal of the Acoustical Society of America. 1947 [accessed 2026 Oct 19];19(1):143–146. http://asa.scitation.org/doi/10.1121/1.1916411. https://doi.org/10.1121/1.1916411

71. Gosling PE. Beginner's Guide to Unit Audio — Part Five: Fitting It into the Room. The Gramophone. 1973 [accessed 2026 Oct 19];(JUNE):127. http://www.gramophonemuseum.com/images/People/Percy%20Wilson/PDF/percy-wilson-gramophone-magazine-1973.pdf

72. Hallum R. Stereophony-a series of perspectives - the stereo listening angle. 2017 [accessed 2026 Oct 31]. https://doi.org/10.13140/rg.2.2.14209.56162

73. Streicher R, Everest FA. The New Stereo Soundbook. 2nd ed. Audio Engineering Associates; 1998.

74. Gerzon MA. Optimal reproduction matrices for multispeaker stereo. In: Audio engineering society convention 91. 1991. http://www.aes.org/e-lib/browse.cfm?elib=5536

75. Griesinger D. Accurate reproduction of binaural recordings through individual headphone equalization and time domain crosstalk cancellation. In: Room Acoustics. Audio Engineering Society; 2019. p 8. https://aes.org/publications/elibrary/elibrary-page/?id=20612

76. Sakamoto N, Gotoh T, Kimura Y. On -Out-of-Head localization- in headphone listening. J. Audio Eng. Soc. 1976 [accessed 2026 Oct 19];24(9):710–716. http://www.aes.org/e-lib/browse.cfm?elib=2591

77. Toole FE. In-Head localization of acoustic images. The Journal of the Acoustical Society of America. 1970 [accessed 2026 Oct 31];48(4B):943–949. https://doi.org/10.1121/1.1912233

78. Mason R. How important is accurate localization in reproduced sound? In: Audio engineering society convention 142. 2017. http://www.aes.org/e-lib/browse.cfm?elib=18635

79. Blauert J et al., editors. The Technology of Binaural Listening. 2013. (Hartmann WM, editor. Modern Acoustics and Signal Processing). https://doi.org/10.1007/978-3-642-37762-4

80. Theile G. Study on the standardisation of studio headphones. EBU Review — Technical. 1983 [accessed 2026 Oct 20];(197). https://hauptmikrofon.de/theile/1983-2_Standardisation-of-studio-headphones_EBU-Rev.1983.pdf

81. International Electrotechnical Commission. IEC 60268-7:2011+A1:2020 Sound system equipment --- Part 7: Headphones and earphones. 2020.

82. International Organization for Standardization. ISO 3741:2010 Acoustics — Determination of sound power levels and sound energy levels of noise sources using sound pressure — Precision methods for reverberation test rooms. CEN; 2010.

83. Olive S, Welti T, McMullin E. Listener preferences for different headphone target response curves. In: AES Convention 134. Audio Engineering Society; 2013. http://www.aes.org/e-lib/browse.cfm?elib=16768

84. Olive SE, Welti T, McMullin E. The Influence of Listeners' Experience, Age, and Culture on Headphone Sound Quality Preferences. In: AES Convention 137. Audio Engineering Society; 2014. p 11. https://aes.org/publications/elibrary/elibrary-page/?id=17500

85. Olive S, Welti T, Khonsaripour O. The preferred low frequency response of in-ear headphones. In: AES Conference: 2016 AES International Conference on Headphone Technology. Audio Engineering Society; 2016. http://www.aes.org/e-lib/browse.cfm?elib=18369

86. Olive S, Welti T, Khonsaripour O. A statistical model that predicts listeners' preference ratings of in-ear headphones: Part 1—Listening test results and acoustic measurements. In: AES Convention 143. Audio Engineering Society; 2017. http://www.aes.org/e-lib/browse.cfm?elib=19237

87. Olive S, Welti T, Khonsaripour O. A statistical model that predicts listeners' preference ratings of in-ear headphones: Part 2—Development and validation of the model. In: AES Convention 143. Audio Engineering Society; 2017. http://www.aes.org/e-lib/browse.cfm?elib=19275

88. Olive S, Khonsaripour O, Welti T. A survey and analysis of consumer and professional headphones based on their objective and subjective performances. In: AES 145th Convention. Audio Engineering Society; 2018. http://www.aes.org/e-lib/browse.cfm?elib=19774

89. Olive S, Welti T, Khonsaripour O. A statistical model that predicts listeners' preference ratings of around-ear and on-ear headphones. In: AES Convention 144. Audio Engineering Society; 2018. http://www.aes.org/e-lib/browse.cfm?elib=19436

90. Olive S, Welti T, Khonsaripour O. Perception & Measurement of Headphone Sound Quality: Is There a Preferred Target Response? 2017 [accessed 2026 Oct 20]. https://www.listeninc.com/wp/media/Perception_and_-Measurement_of_Headphones_Sean_Olive.pdf

91. Olive S, Welti T, McMullin E. Listener preferences for in-room loudspeaker and headphone target responses. In: AES Convention 135. Audio Engineering Society; 2013. http://www.aes.org/e-lib/browse.cfm?elib=17042

92. Toole F. The Measurement and Calibration of Sound Reproducing Systems. Journal of the Audio Engineering Society. 2015 [accessed 2026 Oct 21];63(7/8):512–541. http://www.aes.org/e-lib/browse.cfm?elib=17839. https://doi.org/10.17743/jaes.2015.0064

93. Möller H. Relevant loudspeaker tests in studios in Hi-Fi dealers' demo rooms in the home etc., — using 1/3 octave, pink weighted, random noise. In: 47th Audio Engineering Society Convention. AES; 1974. p 15. https://www.bksv.com/media/doc/17-197.pdf

94. American School of Correspondence, Hulfish DS. Cyclopedia of motion-picture work; a general reference work on the optical lantern, motion head, specific projecting machines, talking pictures, color motography, fixed camera photography, motography, photo-plays, motion-picture theater, management and operation, audience, program, etc. American school of correspondence; 1911. https://archive.org/details/cyclopediaofmoti02amer/mode/2up

95. Jeffress LA. A place theory of sound localization. Journal of Comparative and Physiological Psychology. 1948 [accessed 2026 Oct 31];41(1):35–39. https://doi.org/10.1037/h0061495

96. Pulkki V. Virtual sound source positioning using vector base amplitude panning. J. Audio Eng. Soc. 1997 [accessed 2026 Oct 21];45(6):456–466. http://www.aes.org/e-lib/browse.cfm?elib=7853

97. Blumlein AD. 394,325: Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems. 1933 June 14 [accessed 2026 Oct 21]. https://aes.org/publications/elibrary-page/?id=233

98. Gerzon MA. Applications of Blumlein Shuffing to Stereo Microphone Techniques. In: AES 93rd Convention. 1992. https://aes.org/publications/elibrary-page/?id=6939

99. Eargle J, Streicher R. Acoustical perspectives in commercial two-channel stereophonic recording. In: Audio engineering society conference: 8th international conference: The sound of audio. 1990. http://www.aes.org/e-lib/browse.cfm?elib=5421

100. Ueno Y, Mizumachi M, Horiuchi T. Perceptual optimization of stereo width control methods in headphones. In: AES Convention 150. Audio Engineering Society; 2021. http://www.aes.org/e-lib/browse.cfm?elib=21073

101. International Telecommunication Union. Recommendation ITU-R BS.1534-3 Method for the subjective assessment of intermediate quality level of audio systems. 2015 [accessed 2026 Oct 21]. https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en

102. Schoeffler M et al. webMUSHRA — A Comprehensive Framework for Web-based Listening Tests. Journal of Open Research Software. 2018 [accessed 2026 Oct 21];6(1):8. http://openresearchsoftware.metajnl.com/articles/10.5334/jors.187/. https://doi.org/10.5334/jors.187

103. Bauer BB. Improving headphone listening comfort. J. Audio Eng. Soc. 1965 [accessed 2026 Oct 21];13(4):300–302. http://www.aes.org/e-lib/browse.cfm?elib=1192

104. Linkwitz S. Improved Headphone Listening - Build a stereo-crossfeed circuit. Audio. 1971 [accessed 2026 Oct 21];DECEMBER:42–43. https://worldradiohistory.com/Archive-All-Audio/Archive-Audio/70s/Audio-1971-12.pdf

105. Manor E, Martens W, Marui A, Cabrera D. Nearfield crosstalk increases listener preferences for headphone-reproduced stereophonic imagery. J. Audio Eng. Soc. 2015 [accessed 2026 Oct 21];63(5):324–335. http://www.aes.org/e-lib/browse.cfm?elib=17638. https://doi.org/10.17743/jaes.2015.0022

106. Gilchrest K. Spatial post-processing of hard panned music for headphone reproduction. In: Audio engineering society convention 140. 2016. http://www.aes.org/e-lib/browse.cfm?elib=18291

107. Bech S, Zacharov N. Perceptual Audio Evaluation – Theory, Method and Application. John Wiley & Sons; 2006.

108. Fletcher H. The ear as a measuring instrument. J. Audio Eng. Soc. 1969 [accessed 2026 Oct 21];17(5):532–534. http://www.aes.org/e-lib/browse.cfm?elib=1557

109. Evans KK, Treisman A. Natural cross-modal mappings between visual and auditory features. Journal of Vision. 2010 [accessed 2026 Oct 21];10(1):6–6. https://jov.arvojournals.org/article.aspx?articleid=2121078. https://doi.org/10.1167/10.1.6

110. Roediger R. What Should They Be Called? - APS Observer. Judgement and Decision Making. 2004 [accessed 2026 Oct 21]. http://journal.sjdm.org/roediger.html

111. Ogg M, Slevc LR, Idsardi WJ. The time course of sound category identification: Insights from acoustic features. The Journal of the Acoustical Society of America. 2017 [accessed 2026 Oct 21];142(6):3459–3473. http://asa.scitation.org/doi/10.1121/1.5014057. https://doi.org/10.1121/1.5014057

112. Bizley JK, Cohen YE. The what, where and how of auditory-object perception. Nature Reviews Neuroscience. 2013 [accessed 2026 Oct 21];14(10):693–707. http://www.nature.com/articles/nrn3565. https://doi.org/10.1038/nrn3565

113. Denham SL, Winkler I. Predictive coding in auditory perception: challenges and unresolved questions. European Journal of Neuroscience. 2020 [accessed 2026 Oct 21];51(5):1151–1160. https://onlinelibrary.wiley.com/doi/10.1111/ejn.13802. https://doi.org/10.1111/ejn.13802

114. Yost WA. Fundamentals of hearing: an introduction. Fifth edition. Brill; 2013.

115. Toole FE. Subjective measurements of loudspeaker sound quality and listener performance. J. Audio Eng. Soc. 1985 [accessed 2026 Oct 21];33(1/2):2–32. http://www.aes.org/e-lib/browse.cfm?elib=4465

116. Box GEP, Hunter WG, Hunter JS. Statistics for experimenters: an introduction to design, data analysis, and model building. Wiley; 1978. (Wiley series in probability and mathematical statistics).

117. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology. 2020 [accessed 2026 Oct 22];5(4):536–544. http://www.nature.com/articles/s41564-020-0695-z. https://doi.org/10.1038/s41564-020-0695-z

118. Zentner M, Eerola T. Self-report measures and models. In: Series in affective science. Handbook of music and emotion: Theory, research, applications. Oxford University Press; 2008. p 187–221. https://psycnet.apa.org/record/2010-02543-008

119. Peng ZE et al. Remote testing for psychological and physiological acoustics: Initial report of the P&P Task Force on Remote Testing. 2020. p 050009. http://asa.scitation.org/doi/abs/10.1121/2.0001409. https://doi.org/10.1121/2.0001409

120. International Organization for Standardization. ISO 8586 : 2014 Sensory analysis — General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors. CEN; 2014.

121. Zhao S, Brown CA, Holt LL, Dick F. Robust and efficient online auditory psychophysics with the right auditory hygiene. bioRxiv. 2021 July 19 [accessed 2026 Oct 22]:2021.07.17.452796. https://www.biorxiv.org/content/10.1101/2021.07.17.452796v2. https://doi.org/10.1101/2021.07.17.452796

122. Visser PS, Krosnick JA, Lavrakas PJ. Survey Research. In: Reis HT, Judd CM, editors. Handbook of Research Methods in Social and Personality Psychology. 2nd edition. Cambridge University Press; 2013. https://web.stanford.edu/dept/communication/faculty/krosnick/Survey_Research.pdf

123. Cialdini RB. Littering as an Unobtrusive Measure of Political Attitudes: Messy but Clean. In: Arkin RM, editor. Most underappreciated: 50 prominent social psychologists describe their most unloved work. Oxford University Press; 2011.

124. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008 [accessed 2026 Oct 22];59(4):390–412. https://www.sciencedirect.com/science/article/pii/S0749596X07001398. https://doi.org/10.1016/j.jml.2007.12.005

125. Mortensen CR, Cialdini RB. Full-Cycle Social Psychology for Theory and Application. Social and Personality Psychology Compass. 2010 [accessed 2026 Oct 22];4(1):53–63. http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-9004.2009.00239.x. https://doi.org/10.1111/j.1751-9004.2009.00239.x

126. Paluck EL, Cialdini RB. Field Research Methods. In: Reis HT, Judd CM, editors. Handbook of research methods in social and personality psychology. Second edition. Cambridge University Press; 2014.

127. Popper K. The Logic of Scientific Discovery. 2005.

128. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Springer US; 1993. https://doi.org/10.1007/978-1-4899-4541-9

129. Liu RY, Parelius JM, Singh K. Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference. The Annals of Statistics. 1999 [accessed 2026 Oct 22];27(3):783–840. https://www.jstor.org/stable/120138. https://doi.org/10.1214/aos/1018031260

130. Xu K-M. Using the Bootstrap Method for a Statistical Significance Test of Differences between Summary Histograms. Monthly Weather Review. 2006 [accessed 2026 Oct 22];134(5):1442–1453. http://journals.ametsoc.org/doi/10.1175/MWR3133.1. https://doi.org/10.1175/mwr3133.1

131. Lipshitz SP, Vanderkooy J. The great debate: Subjective evaluation. J. Audio Eng. Soc. 1981 [accessed 2026 Oct 22];29(7/8):482–491. http://www.aes.org/e-lib/browse.cfm?elib=3899

132. International Organization for Standardization. ISO 5495:2005+A1:2016 Sensory analysis — Methodology — Paired comparison test. CEN; 2016.

133. Hermes K, Brookes T, Hummersone C. The influence of dumping bias on timbral clarity ratings. In: Audio engineering society convention 139. 2015. http://www.aes.org/e-lib/browse.cfm?elib=17994

134. Gabrielsson A, Rosenberg U, Sjögren H. Judgments and dimension analyses of perceived sound quality of sound-reproducing systems. The Journal of the Acoustical Society of America. 1974 [accessed 2026 Oct 22];55(4):854–861. http://asa.scitation.org/doi/10.1121/1.1914612. https://doi.org/10.1121/1.1914612

135. Staffeldt H. Differences in the perceived quality of loudspeaker sound reproduction caused by loudspeaker-room-listener interactions. In: Audio engineering society convention 90. 1991. http://www.aes.org/e-lib/browse.cfm?elib=5647

136. Poulton EC. Bias in quantifying judgments. Lawrence Erlbaum; 1989.

137. International Organization for Standardization. ISO 10399:2018 Sensory analysis – Methodology – Duo- trio test. CEN; 2018.

138. Leventhal L, Huynh C-L. Analyzing listening tests with the directional two-tailed test. J. Audio Eng. Soc. 1996 [accessed 2026 Oct 22];44(10):850–863. http://www.aes.org/e-lib/browse.cfm?elib=7881

139. Lawless HT, Heymann H. Sensory Evaluation of Food: Principles and Practices. 2. ed. Springer; 2010. (Food science text series).

140. Brockhoff PB. The statistical power of replications in difference tests. Food Quality and Preference. 2003 [accessed 2026 Oct 22];14(5–6):405–417. https://linkinghub.elsevier.com/retrieve/pii/S095032930300003X. https://doi.org/10.1016/s0950-3293(03)00003-x

141. Leventhal L. Type 1 and type 2 errors in the statistical analysis of listening tests. J. Audio Eng. Soc. 1986 [accessed 2026 Oct 22];34(6):437–453. http://www.aes.org/e-lib/browse.cfm?elib=5265

142. Brockhoff PB, Schlich P. Handling replications in discrimination tests. Food Quality and Preference. 1998 [accessed 2026 Oct 22];9(5):303–312. https://www.sciencedirect.com/science/article/pii/S0950329398000147. https://doi.org/10.1016/s0950-3293(98)00014-7

143. Kunert J, Meyners M. On the triangle test with replications. Food Quality and Preference. 1999 [accessed 2026 Oct 22];10(6):477–482. https://linkinghub.elsevier.com/retrieve/pii/S0950329399000476. https://doi.org/10.1016/s0950-3293(99)00047-6

144. Meyners M, Carr BT, Hasted A. To replicate or not to replicate, or when did we start to ignore the concept of statistical power? Food Quality and Preference. 2020 [accessed 2026 Oct 22];79:103633. https://linkinghub.elsevier.com/retrieve/pii/S0950329318307110. https://doi.org/10.1016/j.foodqual.2019.01.005

145. Kunert J. On repeated difference testing. Food Quality and Preference. 2001 [accessed 2026 Oct 22];12(5–7):385–391. https://linkinghub.elsevier.com/retrieve/pii/S0950329301000295. https://doi.org/10.1016/s0950-3293(01)00029-5

146. van Smeden M et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Medical Research Methodology. 2016 [accessed 2026 Oct 31];16(1):163. https://doi.org/10.1186/s12874-016-0267-3

147. Bech S. Listening Tests on Loudspeakers: A Discussion of Experimental Procedures and Evaluation of the Response Data. In: AES 8th International Conference. AES; 1990. http://www.aes.org/e-lib/browse.cfm?elib=5427

148. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis: with readings. 3rd ed. Macmillan; 1992.

149. Wilcox RR. Introduction to robust estimation and hypothesis testing. 3rd ed. Academic Press; 2012. (Statistical modeling and decision science).

150. Wilcox RR. Understanding and applying basic statistical methods using R. John Wiley & Sons; 2017.

151. International Organization for Standardization. ISO 32000-1:2008 Document management — Portable document format — Part 1: PDF 1.7. ISO; 2008.

152. International Organization for Standardization. ISO 32000-2:2020 Document management — Portable document format — Part 2: PDF 2.0. ISO; 2020.

153. Corporation for Digital Scholarship. Zotero | Your personal research assistant. Zotero. 2021 [accessed 2026 Oct 22]. https://www.zotero.org/

154. Google Books Ngram Viewer: 'listening test', 'listening experiment'. 2021 [accessed 2026 Oct 20]. https://books.google.com/ngrams/graph?content=listening+test%2Clistening+experiment&year_start=1800&year_end=2019&corpus=26&smoothing=3&direct_url=t1%3B%2Clistening%20test%3B%2Cc0%3B.t1%3B%2Clistening%20experiment%3B%2Cc0#t1%3B%2Clistening%20test%3B%2Cc0%3B.t1%3B%2Clistening%20experiment%3B%2Cc0

155. Kahneman D. A perspective on judgment and choice: Mapping bounded rationality. American Psychologist. 2003 [accessed 2026 Oct 31];58(9):697–720. https://doi.org/10.1037/0003-066x.58.9.697

156. Kahneman D. Thinking, fast and slow. Penguin Books; 2012.

157. Tversky A, Kahneman D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review. 1983;90(4):293–315. https://doi.org/10.1037/0033-295x.90.4.293

158. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974 [accessed 2026 Oct 22];185(4157):1124–1131. https://www.jstor.org/stable/1738360. https://doi.org/10.1126/science.185.4157.1124

159. International Organization for Standardization. ISO 13300-2:2006 Sensory analysis: General guidance for the staff of a sensory evaluation laboratory. — Part 2: Recruitment and training of panel leaders. ISO; 2006.

160. International Telecommunication Union. ITU-R BS.1116-3 : Methods for the subjective assessment of small impairments in audio systems. ITU-R; 2015. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf

161. Keppel G. Design and analysis: a researcher's handbook. 2nd ed. Prentice-Hall; 1982.

162. International Telecommunication Union. Recommendation ITU-R BS.1284-2 General methods for the subjective assessment of sound quality. ITU-R; 2019. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1284-2-201901-I!!PDF-E.pdf

163. International Organization for Standardization. ISO 5492:2009 +A1:2017 Sensory analysis— Vocabulary. 2017.

164. International Telecommunication Union. Recommendation ITU-R BS.1387-1 - Method for objective measurements of perceived audio quality. 2001 [accessed 2026 Oct 22]. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1387-1-200111-S!!PDF-E.pdf

165. European Broadcasting Union. Tech 3286 Assessment methods for the subjective evaluation of the quality of sound programme material – Music. EBU; 1997. https://tech.ebu.ch/docs/tech/tech3286.pdf

166. Audio Engineering Society. AES20-1996 r2007; s2008 AES recommended practice for professional audio — Subjective evaluation of loudspeakers. 2008.

167. Legarth SV, Zacharov N. Assessor selection process for multisensory applications. In: Audio engineering society convention 126. 2009. http://www.aes.org/e-lib/browse.cfm?elib=14984

168. Strait D, Kraus N. Playing Music for a Smarter Ear: Cognitive, Perceptual and Neurobiological Evidence. Music Perception. 2011 [accessed 2026 Oct 22];29(2):133–146. https://online.ucpress.edu/mp/article/29/2/133/62520/Playing-Music-for-a-Smarter-Ear-Cognitive. https://doi.org/10.1525/mp.2011.29.2.133

169. Parbery-Clark A, Tierney A, Strait DL, Kraus N. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience. 2012 [accessed 2026 Oct 22];219:111–119. https://linkinghub.elsevier.com/retrieve/pii/S0306452212005106. https://doi.org/10.1016/j.neuroscience.2012.05.042

170. Kraus N, White-Schwoch T. Impact of Life Experiences on Hearing in Noise: The Hearing Journal. 2018 [accessed 2026 Oct 22];71(11):46. http://journals.lww.com/00025572-201811000-00012. https://doi.org/10.1097/01.hj.0000549531.99323.c3

171. Kraus N, White-Schwoch T. Neurobiology of Everyday Communication: What Have We Learned From Music? The Neuroscientist. 2017 [accessed 2026 Oct 22];23(3):287–298. http://journals.sagepub.com/doi/10.1177/1073858416653593. https://doi.org/10.1177/1073858416653593

172. Frankel J, Schwartz J. REAPER. 2021 [accessed 2026 Oct 22]. https://www.reaper.fm/download.php

173. Mateljan I. ARTA Audio Measurement and Analysis Software. 2021 [accessed 2026 Oct 22]. http://www.artalabs.hr/index.htm

174. EASERA - Electronic and Acoustic System Evaluation and Response Analysis. 2024 [accessed 2026 Oct 22]. https://easera.afmg.eu/

175. Bävholm D, Grenander J. Live-Capture Pro. 2019 [accessed 2026 Oct 22]. https://www.wavecapture.eu/LiveCapturePro.html

176. AFMG Technologies GmbH. AFMG SysTune. 2014 [accessed 2026 Oct 22]. https://systune.afmg.eu/

177. Linkwitz S. Shaped tone-burst testing. J. Audio Eng. Soc. 1980 [accessed 2026 Oct 23];28(4):250–258. http://www.aes.org/e-lib/browse.cfm?elib=3989

178. Linkwitz S. Test signals for room reflection and driver distortion investigations. Linkwitz Lab. 2019 [accessed 2026 Oct 23]. http://www.linkwitzlab.com/Ping_tests/Test_signal.htm

179. Zwislocki JJ. Analysis of some auditory characteristics. In: Handbook of Mathematical Psychology : Vol.: 3. Vol. 3. John Wiley and Sons, Inc.; 1965.

180. Mateljan I. ARTA Program for Impulse Response Measurement and Real Time Analysis of Spectrum and Frequency Response - User Manual Version 1.9.4. Artalabs; 2021. https://www.artalabs.hr/download/ARTA-user-manual.pdf

181. Record sales. In: Wikipedia. 2021. https://en.wikipedia.org/w/index.php?title=Record_sales&oldid=1039256573

182. List of best-selling singles. In: Wikipedia. 2021. https://en.wikipedia.org/w/index.php?title=List_of_best-selling_singles&oldid=1040129240

183. List of best-selling albums. In: Wikipedia. 2021. https://en.wikipedia.org/w/index.php?title=List_of_best-selling_albums&oldid=1043486887

184. List of best-selling music artists. In: Wikipedia. 2021. https://en.wikipedia.org/w/index.php?title=List_of_best-selling_music_artists&oldid=1043590737

185. Wittek H, Theile G. The recording angle - based on localisation curves. In: Audio engineering society convention 112. 2002. http://www.aes.org/e-lib/browse.cfm?elib=11307

186. Sengpiel E. Calculation of the direction of phantom sources by interchannel level difference and time difference stereophony stereo time of arrival - localization curves - sengpielaudio Sengpiel Berlin. Tontechnik-Rechner - sengpielaudio. 2021 June 12 [accessed 2026 Oct 23]. http://www.sengpielaudio.com/calculator-localisationcurves.htm

187. Benjamin E. An experimental verification of localization in two-channel stereo. In: AES Convention 121. Audio Engineering Society; 2006. http://www.aes.org/e-lib/browse.cfm?elib=13802

188. Ciba S, Wlodarski A, Maempel H-J. WhisPER – a new tool for performing listening tests. In: AES Convention 126. Audio Engineering Society; 2009. http://www.aes.org/e-lib/browse.cfm?elib=14945

189. Giner AV. Scale - A Software Tool for Listening Experiments. 2013:4.

190. De Man B, Reiss JD. APE: Audio perceptual evaluation toolbox for MATLAB. In: AES Convention 136. Audio Engineering Society; 2014. http://www.aes.org/e-lib/browse.cfm?elib=17160

191. Kraft S, Zolzer U. BeaqleJS: HTML5 and JavaScript based Framework for the Subjective Evaluation of Audio Quality. 2014. https://github.com/HSU-ANT/beaqlejs

192. Jillings N et al. Web Audio Evaluation Tool: A framework for subjective assessment of audio. 2016 [accessed 2026 Oct 23]. https://www.open-access.bcu.ac.uk/3272/1/web%20audio.pdf

193. Gozynski. APL-Huddersfield/BO-LTS: BO-LTS v1.0. 2020 [accessed 2026 Oct 31]. https://doi.org/10.5281/zenodo.4145373

194. Johnson D, Lee H. HULTI-GEN v2.0.5. 2020 [accessed 2026 Oct 31]. https://doi.org/10.5281/zenodo.4302080

195. Rumsey F. Listening tests and perception research. J. Audio Eng. Soc. 2021 [accessed 2026 Oct 23];69(6):433–438. http://www.aes.org/e-lib/browse.cfm?elib=21111

196. Stecker GC et al. Remote testing for psychological and physiological acoustics: Initial report of the ASA P&P Task Force on Remote Testing. The Journal of the Acoustical Society of America. 2020 [accessed 2026 Oct 23];148(4):2713–2713. https://doi.org/10.1121/1.5147519

197. W3C Audio Working Group. Web Audio API. 2021 [accessed 2026 Oct 23]. https://www.w3.org/TR/webaudio-1.0/

198. Rumsey F. The Importance of Loudness. Journal of the Audio Engineering Society. 2021 [accessed 2026 Oct 23];69(3):211–213. https://aes.org/publications/elibrary/elibrary-page/?id=21029

199. European Broadcasting Union. R 128: Loudness Normalisation and Permitted Maximum Level of Audio Signals. EBU; 2020. https://tech.ebu.ch/docs/r/r128.pdf

200. International Telecommunication Union. Recommendation ITU-R BS.1770-4 Algorithms to measure audio programme loudness and true-peak audio level. 2015 [accessed 2026 Oct 23]. https://www.itu.int/rec/R-REC-BS.1770/en

201. European Broadcasting Union. R 128 s1: Loudness Parameters for Short-Form Content (Adverts, Promos, etc.). 2020 [accessed 2026 Oct 18]. https://tech.ebu.ch/docs/r/r128s1.pdf

202. RME Audio. Digicheck. 2020 [accessed 2026 Oct 23]. https://www.rme-audio.de/digicheck.html

203. European Broadcasting Union. Tech 3341-2016 Loudness Metering: 'EBU Mode' Metering to Supplement EBU R 128 Loudness Normalization. EBU; 2016. https://tech.ebu.ch/docs/tech/tech3341.pdf

204. Mason A. WHP 202 Terminology for loudness and level dBTP, LU, and all that. 2011 [accessed 2026 Oct 22]. http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP202.pdf

205. International Telecommunication Union. P.800 Methods for objective and subjective assessment of quality. ITU-T; 1996. https://www.itu.int/rec/T-REC-P.800-199608-I

206. Bregman AS, Ahad PA. Demonstrations to accompany Bregman's: Auditory Scene Analysis - The perceptual organization of sound. MIT Press; 1990. http://www.rctn.org/bruno/data/auditory_demonstrations/ASA-Demo-Booklet9V4.pdf

207. mcdermottLab. mcdermottLab/HeadphoneCheck. 2021 [accessed 2026 Oct 23]. https://github.com/mcdermottLab/HeadphoneCheck

208. Linkwitz S. Phantom stereo-center-speaker. Linkwitz Lab. 2019 [accessed 2026 Oct 23]. https://www.linkwitzlab.com/Loudspeaker-Room/phantom_center.htm

209. Ennis DM, Bi J. The Beta-Binomial Model: Accounting for Inter-Trial Variation in Replicated Difference and Preference Tests. Journal of Sensory Studies. 1998 [accessed 2026 Oct 23];13(4):389–412. http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-459X.1998.tb00097.x. https://doi.org/10.1111/j.1745-459x.1998.tb00097.x

210. Jones LV, Tukey JW. A Sensible Formulation of the Significance Test. 2000:4. https://doi.org/10.1037/1082-989x.5.4.411

211. Cumming G, Maillardet R. Confidence intervals and replication: Where will the next mean fall? Psychological Methods. 2006 [accessed 2026 Oct 31];11(3):217–227. https://doi.org/10.1037/1082-989x.11.3.217

212. Erceg-Hurn D, Mirosevich V. Modern Robust Statistical Methods An Easy Way to Maximize the Accuracy and Power of Your Research. The American psychologist. 2008;63:591–601. https://doi.org/10.1037/0003-066x.63.7.591

213. Silva D, Clarke P. Some Initiatives on Combining Data to Support Small Area Statistics and Analytical Requirements at ONS-UK. IAOS 2008 Conference on Reshaping Official Statistics. 2008 [accessed 2026 Oct 23];1. https://www.researchgate.net/profile/Denise-Silva-18/publication/228391798_Some_Initiatives_on_Combining_Data_to_Support_Small_Area_Statistics_and_Analytical_Requirements_at_ONS-UK/links/02e7e5395254bb8446000000/Some-Initiatives-on-Combining-Data-to-Support-Small-Area-Statistics-and-Analytical-Requirements-at-ONS-UK.pdf

214. Lawless HT, Heymann H. Discrimination Testing. In: Sensory Evaluation of Food: Principles and Practices. Springer; 2013.

215. Field AP, Wilcox RR. Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. Behaviour Research and Therapy. 2017 [accessed 2026 Oct 23];98:19–38. https://www.sciencedirect.com/science/article/pii/S0005796717301067. https://doi.org/10.1016/j.brat.2017.05.013

216. Wilcox RR, Serang S. Hypothesis Testing, p Values, Confidence Intervals, Measures of Effect Size, and Bayesian Methods in Light of Modern Robust Techniques. Educational and Psychological Measurement. 2017 [accessed 2026 Oct 23];77(4):673–689. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5991786/. https://doi.org/10.1177/0013164416667983

217. Wilcox R, Rousselet G. A guide to robust statistical methods: Illustrations using R. 2017 [accessed 2026 Oct 23]:247268 Bytes. https://figshare.com/articles/journal_contribution/A_guide_to_robust_statistical_methods_Illustrations_using_R/5047585/1. https://doi.org/10.6084/m9.figshare.5047585.v1

218. Wilcox R. Identifying Which of J Independent Binomial Distributions Has the Largest Probability of Success. Journal of Modern Applied Statistical Methods. 2020 [accessed 2026 Oct 23];18(2):2–9. https://digitalcommons.wayne.edu/jmasm/vol18/iss2/5. https://doi.org/10.22237/jmasm/1604190960

219. Babbage C. Reflections on the decline of science in England, and on some of its causes. B. Fellowes; 1830. (Reflections on the decline of science in england: And on some of its causes, by charles babbage (1830). To which is added on the alleged decline of science in england, by a foreigner (gerard moll) with a foreword by michael faraday (1831)). https://books.google.co.uk/books?id=oXbRotQLLtsC

220. Jennings RC. Data selection and responsible conduct: Was millikan a fraud? Science and Engineering Ethics. 2004 [accessed 2026 Oct 23];10(4):639–653. http://link.springer.com/10.1007/s11948-004-0044-2. https://doi.org/10.1007/s11948-004-0044-2

221. Gelman A, Loken E. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no ‘fishing expedition’ or ‘p-hacking’ and the research hypothesis was posited ahead of time. 2014 Nov 14 [accessed 2026 Oct 23]:17. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

222. R Core Team. R: A language and environment for statistical computing. 2021 [accessed 2026 Oct 23]. https://www.R-project.org/

223. Shafranovich Y. Common Format and MIME Type for Comma-Separated Values (CSV) Files. Internet Engineering Task Force; 2005. Report No.: RFC 4180. https://datatracker.ietf.org/doc/rfc4180. https://doi.org/10.17487/RFC4180

224. Lex A et al. UpSet: Visualization of Intersecting Sets. IEEE Transactions on Visualization and Computer Graphics. 2014;20(12):1983–1992. https://doi.org/10.1109/tvcg.2014.2346248

225. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties Hancock J, editor. Bioinformatics. 2017 [accessed 2026 Oct 20];33(18):2938–2940. https://academic.oup.com/bioinformatics/article/33/18/2938/3884387. https://doi.org/10.1093/bioinformatics/btx364

226. Schlich P. Risk tables for discrimination tests. Food Quality and Preference. 1993;4(3):141–151. https://doi.org/10.1016/0950-3293(93)90157-2

227. Wilcox R. A Note on Inferences About the Probability of Success. Journal of Modern Applied Statistical Methods. 2020 [accessed 2026 Oct 24];18(1):jmasm.eP3296. http://digitalcommons.wayne.edu/jmasm/vol18/iss1/30. https://doi.org/10.22237/jmasm/1556670420

228. Schilling MF, Doi JA. A Coverage Probability Approach to Finding an Optimal Binomial Confidence Procedure. The American Statistician. 2014 [accessed 2026 Oct 24];68(3):133–145. http://www.tandfonline.com/doi/abs/10.1080/00031305.2014.899274. https://doi.org/10.1080/00031305.2014.899274

229. Agresti A, Coull BA. Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions. The American Statistician. 1998 [accessed 2026 Apr 22];52(2):119–126. http://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480550. https://doi.org/10.1080/00031305.1998.10480550

230. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association. 1927 [accessed 2026 Oct 24];22(158):209–212. https://www.jstor.org/stable/2276774. https://doi.org/10.2307/2276774

231. Zou GY, Huang W, Zhang X. A note on confidence interval estimation for a linear function of binomial proportions. Computational Statistics & Data Analysis. 2009 [accessed 2026 Oct 24];53(4):1080–1085. https://linkinghub.elsevier.com/retrieve/pii/S0167947308004623. https://doi.org/10.1016/j.csda.2008.09.033

232. Altman D, Machin D, Bryant T, Gardner M, editors. Statistics with confidence: Confidence intervals and statistical guidelines. Wiley; 2013.

233. Matsumoto M. Mersenne Twister: A random number generator (since 1997/10). 2021 Aug 1 [accessed 2026 Oct 24]. http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/emt.html

234. Schlich P, Dacremont C, Brockhoff PB. Application of replicated difference testing. Food Quality and Preference. 2000:4. https://doi.org/10.1016/s0950-3293(99)00037-3

235. Student. The Probable Error of a Mean. Biometrika. 1908 [accessed 2026 Oct 24];6(1):1–25. https://www.jstor.org/stable/2331554. https://doi.org/10.2307/2331554

236. Rudmose W. The case of the missing 6 dB. The Journal of the Acoustical Society of America. 1982 [accessed 2026 Oct 24];71(3):650–659. http://asa.scitation.org/doi/10.1121/1.387540. https://doi.org/10.1121/1.387540

237. Munson WA, Wiener FM. In Search of the Missing 6 Db. The Journal of the Acoustical Society of America. 1952 [accessed 2026 Oct 31];24(4):449–449. https://doi.org/10.1121/1.1917482

238. Morgan DE, Dirks DD. Loudness discomfort level under earphone and in the free field: the effects of calibration methods. The Journal of the Acoustical Society of America. 1974 [accessed 2026 Oct 31];56(1):172–178. https://doi.org/10.1121/1.1903249

239. Killion MC. Revised estimate of minimum audible pressure: Where is the ' 'missing 6 dB' '? The Journal of the Acoustical Society of America. 1978 [accessed 2026 Oct 24];63(5):1501–1508. http://asa.scitation.org/doi/10.1121/1.381844. https://doi.org/10.1121/1.381844

240. Keidser G, Katsch R, Dillon H, Grant F. Relative loudness perception of low and high frequency sounds in the open and occluded ear. The Journal of the Acoustical Society of America. 2000 [accessed 2026 Oct 31];107(6):3351–3357. https://doi.org/10.1121/1.429406

241. Völk F, Fastl H. Locating the missing 6 dB by loudness calibration of binaural synthesis. In: AES Convention 131. Audio Engineering Society; 2011. http://www.aes.org/e-lib/browse.cfm?elib=16014

242. Brinkmann F, Lindau A, Weinzierl S. On the authenticity of individual dynamic binaural synthesis. The Journal of the Acoustical Society of America. 2017 [accessed 2026 Oct 31];142(4):1784–1795. https://doi.org/10.1121/1.5005606

243. Denk F et al. The 'Missing 6 dB' Revisited: Influence of Room Acoustics and Binaural Parameters on the Loudness Mismatch Between Headphones and Loudspeakers. Frontiers in Psychology. 2021 [accessed 2026 Oct 24];12:623670. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8032957/. https://doi.org/10.3389/fpsyg.2021.623670

244. Beranek LL. Acoustic measurements. John Wiley and Sons, Inc.; 1949.

245. Mathers CD, Lansdowne KFL. Hearing risk to wearers of circumaural head- phones: an investigation. British Broadcasting Corporation - Research Department; 1979. Report No.: BBC RD 1979/3. http://downloads.bbc.co.uk/rd/pubs/reports/1979-03.pdf

246. Zahorik P, Wightman FL. Loudness constancy with varying sound source distance. Nature Neuroscience. 2001 [accessed 2026 Oct 24];4(1):78–83. http://www.nature.com/articles/nn0101_78. https://doi.org/10.1038/82931

247. Sivonen VP, Ellermeier W. Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. The Journal of the Acoustical Society of America. 2006 [accessed 2026 Oct 31];119(5):2965–2980. https://doi.org/10.1121/1.2184268

248. Dudley H. Fundamentals of speech synthesis. J. Audio Eng. Soc. 1955 [accessed 2026 Oct 24];3(4):170–185. http://www.aes.org/e-lib/browse.cfm?elib=9

249. Perez-Ortiz M, Mantiuk RK. A practical guide and software for analysing pairwise comparison experiments. arXiv:1712.03686 [cs, stat]. 2017 Dec 15 [accessed 2026 Oct 24]. http://arxiv.org/abs/1712.03686

250. Leijon A, Dahlquist M, Smeds K. Bayesian analysis of paired-comparison sound quality ratings. The Journal of the Acoustical Society of America. 2019 [accessed 2026 Oct 24];146(5):3174–3183. http://asa.scitation.org/doi/10.1121/1.5131024. https://doi.org/10.1121/1.5131024

251. Gantt chart. In: Wikipedia. 2023. https://en.wikipedia.org/w/index.php?title=Gantt_chart&oldid=1139942621

252. Serra-Garcia M, Gneezy U. Nonreplicable publications are cited more than replicable ones. Science Advances. 2021 [accessed 2026 Oct 24];7(21):eabd1705. https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.abd1705. https://doi.org/10.1126/sciadv.abd1705

253. Mok BA et al. Web-based Psychoacoustics: Hearing Screening, Infrastructure, and Validation. bioRxiv. 2021 May 11 [accessed 2026 Oct 24]:2021.05.10.443520. https://www.biorxiv.org/content/10.1101/2021.05.10.443520v1. https://doi.org/10.1101/2021.05.10.443520

254. Kleiner M. Electroacoustics. CRC Press; 2013.

255. Michel J-B et al. Quantitative Analysis of Culture Using Millions of Digitized Books. Science. 2011 [accessed 2026 Oct 20];331(6014):176–182. https://www.science.org/doi/10.1126/science.1199644. https://doi.org/10.1126/science.1199644

256. Ilkless. A Deep Dive Into Harman Curves - The Long Way Round (Part 3) - Headphonesty. 2021 [accessed 2026 Oct 20]. https://www.headphonesty.com/2020/04/harman-target-curves-part-3/

257. Plenge G. On the differences between localization and lateralization. The Journal of the Acoustical Society of America. 1974 [accessed 2026 Oct 24];56(3):944–951. http://asa.scitation.org/doi/10.1121/1.1903353. https://doi.org/10.1121/1.1903353

List of Tables

3.1 Sensory Assessor Types[120,163] . . . 23
3.2 Synthetic Signal and Programme Excerpts for Polarity Experiment . . . 25
3.3 Programme Excerpts for Crosstalk Experiment . . . 27
3.3 Programme Excerpts for Crosstalk Experiment . . . 28
4.1 Summary of Sample Ages (years) . . . 33
4.2 Sample Groups for Cluster Analysis . . . 34
4.3 Polarity Experiment Duo-Trio Test: Results Summary . . . 35
4.4 Crosstalk Experiment Paired Comparison Test: Results Summary . . . 35
7.1 *POL3* Polarity Experiment: Full Results . . . 67
7.2 *POL4* Polarity Experiment: Full Results . . . 68
7.3 *POL8* Polarity Experiment: Full Results . . . 69
7.4 *MP6* Crosstalk Experiment: Full Results . . . 70
7.5 *MP12* Crosstalk Experiment: Full Results . . . 71
7.6 *MP18* Crosstalk Experiment: Full Results . . . 72
7.7 *MP22* Crosstalk Experiment: Full Results . . . 73
7.8 *MP47* Crosstalk Experiment: Full Results . . . 74
7.9 Questionnaire . . . 75
7.10 Difference of the Difference Levels of LR outputs for MS and PAN Crosstalk Methods. . . . 78

List of Figures

2.1 'Record—Playback of acoustic waveform showing that a compression at the microphone should be reproduced as a compression at the loudspeaker'.[12(fig1)] Reproduced from Greiner RA, Melton D; 1994. 5

2.2 Two loudspeakers reproducing the same programme content creating equal net phasors, below 700 Hz, at both ears – reproduced from Eargle J, 2004.[61(figs11–2)] The equal phasors result in a phantom centre image in front of a forward-facing listener. L_L is the signal vector at the Left ear from Left loudspeaker, L_R is the crosstalk signal vector at Left ear from Right loudspeaker. L_T is vector addition of L_L and L_R, a forward phasor for partial contribution to a phantom centre image. Adapted from Bock TM, Keele, Jr. DB (Don); 1986.[62] 9

2.3 Reproduced from Bauer BB, 1961,[69(fig3)] a graph prediction of relative responses at the left (L) and right(R) ears from a loudspeaker positioned at 45° horizontal orientation, derived from Wiener[70]. The "head-on" function (L,R, 0°) is shown as relatively flat. Note between the ear dimension as 8 inches (203 mm), and approximated increased path-length to right ear of 5.5 inches (140 mm). It may be assumed that the common recommended stereo loudspeaker arrangement of ±30° had yet to be realised until a decade or more later.[71(p127),72] 11

2.4 Stereophonic amplitude panning formulated with vectors: $\mathbf{p = g_1\, l_1 + g_2\, l_2}$ where $\mathbf{l_1\ l_2}$ are loudspeaker vectors; $\mathbf{g_1\ g_2}$ are the gains or amplitude levels for loudspeaker **channel 1** and **2**. Altering the relative gains alters the virtual source position **p** on the active arc. Reproduced from Pulkki V, 1997.[96(fig3)] 13

2.5 Schematic of a width control for amplitude stereophony. The "sum" channels are here denoted as MID signals **M** and **M'**, the "difference" channels denoted as **S** and **S'**. Reproduced from Gerzon MA, 1992.[98(fig2)] . 14

2.6 A physical sound field stimulates the ear, resulting in neural responses that combine to form perception of the auditory event. The frequency, intensity and time attributes of sound are first coded by the peripheral auditory nervous system. This neural code is processed by higher neural centres to allow the listener to determine the sources of sound. This neural information is integrated with other sensory information, and that of experience, leading to behavioural responses. Reproduced from Yost WA, 2013.[114(fig1.2)] 16
3.1 *POL3* test signal, a 300 Hz Gauss envelope tone burst, 4 sine periods. Zoom 1:8 26
3.2 *POL3* test signal, a 300 Hz Gauss envelope tone burst, 4 sine periods. Zoom 2:1 27
3.3 'ILDs for a [typical] listener and 60° stereo triangle'.[187(fig19)] Reproduced from Benjamin E, 2006. 29
7.1 Number Correct versus Estimated Discriminators. Reproduced from Lawless HT, Heymann H; 2010.[139(p107)] 76
7.2 Summary of assessor categories employed in sensory analysis, as defined in ISO standard 8586-2,[120] applied to the food industry and recommended for adoption in the field of audio. Reproduced from Legarth SV, 2009.[167(p4)] 77
7.3 Diagrammatic Horizontal Sections Showing Types of Earphones. Reproduced from Kleiner M, 2013.[254(fig24.1)] 79
7.4 A graph showing how the phrases 'listening test' and 'listening experiment' have occurred in the *English 2019* corpus of books from 1900 to 2019 (*Google Books Ngram Viewer: 'listening test', 'listening experiment'*, 2021).[154,255] 80
7.5 Plot showing Participants' Age Distribution (Years) 81
7.6 *UpSet*[225] Graph #1, showing Relationships Between Subject Groups . . . 82
7.7 *UpSet*[225] graph #2, showing Relationships Between Subject Groups 83
7.8 Earphone #1 phase estimate (note Phase range) 84
7.9 Earphone #2 phase estimate 84
7.10 Earphone #3 phase estimate 85
7.11 Earphone #4 phase estimate 85
7.12 Earphone #5 phase estimate (note Phase range) 86
7.13 Earphone #6 phase estimate (note Phase range) 86
7.14 Earphone #7 phase estimate (note Phase range) 87
7.15 Earphone #8 phase estimate 87
7.16 Earphone #9 phase estimate 88
7.17 Earphone #10 phase estimate 88
7.18 Earphone #11 phase estimate 89

7.19 This graph compares the Harman curves published over the years, showing substantial variation in the bass and treble. Notably, the in-ear target [IE] reflects the difference in a blocked ear canal makes, as the blocked canal has it[s] own sets of resonances that the IEM needs to compensate for. (From: Harman). Note: OE — Over the Ear headphone; IE — In-Ear earphone. Reproduced from Ilkless, 2021.[256] 89
7.20 'Optimum curve for HIFI equipment. Measured in the actual listening room'.[93(fig5)] Reproduced from Möller H, 1974. 90
7.21 Simulation of the directional characteristic of the human head related to the sound incident direction (30°). Reproduced from Plenge G, 1974.[257(fig8)] 91
7.22 Direction of incidence of sound for sinusoids and Gaussian tone bursts of critical bandwidth so that the same direction of the auditory event is achieved as with broad band noise from the directions $\delta = 0$', 40°, 320°. Solid lines: sinusoidal signals (after Sandel et al. 1955 sound pressure level 35 dB, 5 subjects, head immobilized). Dashed lines: Gaussian tone bursts (after Boerger 1965a ,b ;50-60 phon, 4–7 subjects, no head movement). Reproduced from Blauert J, 1996.[63(fig2.3)] 92
7.23 *webMUSHRA* 'paired comparison example' page, personal computer screenview. 93
7.24 *webMUSHRA* 'paired comparison example' page, smartphone screenview. 93
7.25 *webMUSHRA* 'duo-trio example' page,personal computer screenview. . . 94
7.26 *webMUSHRA* 'duo-trio example' page, smartphone screenview. 94
7.27 *webMUSHRA* 'Level Calibration' page, personal computer screenview . . 95
7.28 The basic configuration of a Mid/Side stereo pickup; the mid microphone is aimed at the 0° axis. The side microphone is a bidirectional, with its null plane aligned with the principal axis of the mid microphone. By convention, the positive lobe of the side microphone is oriented to the left side of the stereo soundstage. Reproduced from Streicher R, Everest FA; 1998.[73(fig7.17)] 96
7.29 Example REAPER mix view showing relative crosstalk level for 1 kHz Sine routed to Left channel only of MS and PAN group busses.[172] Metered at PAN bus output using RME Digicheck.[202] 97
7.30 Number of assessors needed for a duo-trio test. Adapted from Schlich P, 1993.[226] 100

7 APPENDIX

7.1 Supplementary Tables

Table 7.1: *POL3* Polarity Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.586	0.025	0.95	(0.536, 0.634)
all	96	384	0.586	0.025	0.99	(0.520, 0.649)
all	96	384	0.586	0.025	0.999	(0.502, 0.665)
age $<$ 48	48	192	0.563	0.035	0.95	(0.492, 0.631)
age $\geq$ 48	48	192	0.609	0.035	0.95	(0.539,0.676)
intersect	24	96	0.615	0.049	0.95	(0.515,0.706)
!intersect	72	288	0.576	0.029	0.95	(0.519, 0.632)
closed 'phone	55	220	0.609	0.033	0.95	(0.543, 0.671)
!closed 'phone	41	164	0.555	0.038	0.95	(0.478, 0.629)
multilingual	48	192	0.604	0.035	0.95	(0.534, 0.671)
multilingual	48	192	0.604	0.035	0.99	(0.511, 0.690)
listen test experience	56	224	0.567	0.033	0.95	(0.501, 0.630)
musical training	64	256	0.598	0.030	0.95	(0.537, 0.656)
musical training	64	256	0.598	0.030	0.99	(0.517, 0.673)

Table 7.2: *POL4* Polarity Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.568	0.025	0.95	(0.518, 0.616)
all	96	384	0.568	0.025	0.99	(0.502, 0.631)
age $<$ 48	48	192	0.536	0.036	0.95	(0.466, 0.606)
age $\geq$ 48	48	192	0.599	0.035	0.95	(0.528, 0.666)
age $\geq$ 48	48	192	0.599	0.035	0.99	(0.506, 0.685)
intersect	24	96	0.531	0.050	0.95	(0.432, 0.628)
!intersect	72	288	0.580	0.029	0.95	(0.522, 0.635)
closed 'phone	55	220	0.564	0.033	0.95	(0.498, 0.628)
!closed 'phone	41	164	0.573	0.038	0.95	(0.497, 0.646)
multilingual	48	192	0.536	0.036	0.95	(0.466, 0.606)
listen test experience	56	224	0.571	0.033	0.95	(0.506, 0.635)
musical training	64	256	0.551	0.031	0.95	(0.490, 0.611)

Table 7.3: *POL8* Polarity Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.547	0.025	0.95	(0.497, 0.596)
all	96	384	0.547	0.025	0.90	(0.505, 0.588)
age $<$ 48	48	192	0.563	0.035	0.95	(0.492, 0.631)
age $\geq$ 48	48	192	0.531	0.036	0.95	(0.461, 0.601)
intersect	24	96	0.448	0.050	0.95	(0.352, 0.548)
!intersect	72	288	0.580	0.029	0.95	(0.522, 0.635)
!intersect	72	288	0.580	0.029	0.99	(0.504, 0.652)
closed 'phone	55	220	0.577	0.033	0.95	(0.511, 0.641)
!closed 'phone	41	164	0.506	0.039	0.95	(0.430, 0.582)
multilingual	48	192	0.521	0.036	0.95	(0.450, 0.590)
listen test experience	56	224	0.496	0.033	0.95	(0.431, 0.561)
musical training	64	256	0.520	0.031	0.95	(0.458, 0.580)

Table 7.4: *MP6* Crosstalk Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.469	0.025	0.95	(0.419, 0.519)
age <48	48	192	0.464	0.036	0.95	(0.394, 0.534)
age ≥ 48	48	192	0.474	0.036	0.95	(0.405, 0.544)
intersect	24	96	0.531	0.050	0.95	(0.432, 0.628)
!intersect	72	288	0.448	0.029	0.95	(0.392, 0.506)
!intersect	72	288	0.448	0.029	0.90	(0.400, 0.496)
closed 'phone	55	220	0.468	0.033	0.95	(0.403, 0.534)
!closed 'phone	41	164	0.470	0.039	0.95	(0.395, 0.546)
multilingual	48	192	0.464	0.036	0.95	(0.394, 0.534)
listen test experience	56	224	0.455	0.033	0.95	(0.391, 0.521)
musical training	64	256	0.465	0.031	0.95	(0.405, 0.526)

Table 7.5: *MP12* Crosstalk Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.477	0.025	0.95	(0.427, 0.527)
age $<$ 48	48	192	0.510	0.036	0.95	(0.440, 0.580)
age $\geq$ 48	48	192	0.443	0.036	0.95	(0.374, 0.513)
age $\geq$ 48	48	192	0.443	0.036	0.90	(0.385, 0.502)
intersect	24	96	0.406	0.049	0.95	(0.313, 0.506)
intersect	24	96	0.406	0.049	0.90	(0.327, 0.490)
!intersect	72	288	0.500	0.029	0.95	(0.443, 0.557)
closed 'phone	55	220	0.468	0.033	0.95	(0.403, 0.534)
!closed 'phone	41	164	0.488	0.039	0.95	(0.412, 0.564)
multilingual	48	192	0.453	0.036	0.95	(0.384, 0.524)
listen test experience	56	224	0.446	0.033	0.95	(0.383, 0.512)
listen test experience	56	224	0.446	0.033	0.90	(0.393, 0.501)
musical training	64	256	0.469	0.031	0.95	(0.409, 0.530)

Table 7.6: *MP18* Crosstalk Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.495	0.025	0.95	(0.445, 0.545)
age <48	48	192	0.500	0.036	0.95	(0.430, 0.570)
age ≥ 48	48	192	0.490	0.036	0.95	(0.420, 0.560)
intersect	24	96	0.469	0.050	0.95	(0.372, 0.568)
!intersect	72	288	0.503	0.029	0.95	(0.446, 0.561)
closed 'phone	55	220	0.500	0.033	0.95	(0.434, 0.566)
!closed 'phone	41	164	0.488	0.039	0.95	(0.412, 0.564)
multilingual	48	192	0.505	0.036	0.95	(0.435, 0.575)
listen test experience	56	224	0.509	0.033	0.95	(0.444, 0.574)
musical training	64	256	0.496	0.031	0.95	(0.435, 0.557)

Table 7.7: *MP22* Crosstalk Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.508	0.025	0.95	(0.458, 0.557)
age <48	48	192	0.536	0.036	0.95	(0.466, 0.606)
age ≥ 48	48	192	0.479	0.036	0.95	(0.410, 0.550)
intersect	24	96	0.458	0.050	0.95	(0.362, 0.558)
!intersect	72	288	0.524	0.029	0.95	(0.467, 0.581)
closed 'phone	55	220	0.523	0.033	0.95	(0.457, 0.588)
!closed 'phone	41	164	0.488	0.039	0.95	(0.412, 0.564)
multilingual	48	192	0.484	0.036	0.95	(0.415, 0.555)
listen test experience	56	224	0.509	0.033	0.95	(0.444, 0.574)
musical training	64	256	0.492	0.031	0.95	(0.432, 0.553)

Table 7.8: *MP47* Crosstalk Experiment: Full Results

GROUP	N	n	$\hat{p}$	**SE**	$1-\alpha$	**CI**
all	96	384	0.544	0.025	0.95	(0.494, 0.593)
all	96	384	0.544	0.025	0.90	(0.502, 0.586)
age $<$ 48	48	192	0.563	0.035	0.95	(0.492, 0.631)
age $<$ 48	48	192	0.563	0.035	0.90	(0.503, 0.620)
age $\geq$ 48	48	192	0.526	0.036	0.95	(0.456, 0.595)
intersect	24	96	0.563	0.050	0.95	(0.463, 0.657)
!intersect	72	288	0.538	0.029	0.95	(0.480, 0.595)
closed 'phone	55	220	0.555	0.033	0.95	(0.488, 0.619)
!closed 'phone	41	164	0.530	0.039	0.95	(0.454, 0.605)
multilingual	48	192	0.552	0.036	0.95	(0.481, 0.621)
listen test experience	56	224	0.558	0.033	0.95	(0.493, 0.622)
listen test experience	56	224	0.558	0.033	0.90	(0.503, 0.612)
musical training	64	256	0.547	0.031	0.95	(0.486, 0.607)

Table 7.9: Questionnaire

QUESTION	RESPONSE CHOICE
“What is your sex?”	Female ; Male
“Did you adjust the volume level of your device after the calibration?”	No ; Yes
“What type of headphones are you using?”	Closed-back ; Open-back ; In-Ear ; Moulds
“Do you have any formal musical training / play and instrument / sing?”	No ; Yes
“Can you speak more than one language?”	No ; Yes
“Have you taken part in a listening experiment before?”	No ; Yes
“Do you have a background in professional audio (e.g. sound / recording engineer, researcher)?”	No ; Yes
Feedback	Text box
Age	Text box
“Your eMail”	Text box

N	Minimum correct, $p = 1/2$	Estimated number discriminating	Minimum correct, $p = 1/3$	Estimated number discriminating
10	9	6	7	4
15	12	7	9	5
20	15	8	11	6
25	18	9	13	7
30	20	10	15	7
35	23	11	17	8
40	26	11	19	8
45	29	12	21	9
50	32	13	23	9
55	35	13	25	9
60	37	14	27	10
65	40	14	29	10
70	43	15	31	10
75	46	15	33	11
80	49	16	35	11
85	51	16	36	11
90	54	17	39	12
95	57	17	40	12
100	59	17	42	12

Minimum correct gives the required number for significance at $p < 0.05$, one tailed

Figure 7.1: Number Correct versus Estimated Discriminators. Reproduced from Lawless HT, Heymann H; 2010.[139(p107)]

Assessor category	Definition
Assessor	Any person taking part in a sensory test
Naive assessor	A person who does not meet any particular criterion
Initiated assessor	A person who has already participated in a sensory test
Selected assessor	Assessor chosen for his/her ability to carry out a sensory test
Expert	In the general sense, a person who through knowledge or experience has competence to give an opinion in the field about which he/she is consulted. (Please note that the term *expert* does not provide any indication regarding the qualification or suitability of the individual to perform listening tests.)
Expert assessor	Selected assessor with a high degree of sensory sensitivity and experience in sensory methodology, who is able to make consistent and repeatable sensory assessments of various products
Specialised expert assessor	Expert assessor who has additional experience as a specialist in the product and/or process and/or marketing, and who is able to perform sensory analysis of the product and evaluate or predict effects of variations relating to raw materials, recipes, processing, storage, ageing, and so on.

Figure 7.2: Summary of assessor categories employed in sensory analysis, as defined in ISO standard 8586-2,[120] applied to the food industry and recommended for adoption in the field of audio. Reproduced from Legarth SV, 2009.[167(p4)]

Table 7.10: Difference of the Difference Levels of LR outputs for MS and PAN Crosstalk Methods.

TITLE	**ARTIST / GROUP**	***diff* MS – *diff* PAN**
It's Too Late	Carole King	–60.6 dB Peak –63.5 dB RMS-I
What'd I Say (Parts 1 And 2)	Ray Charles	–56.3 dB Peak –57 dB RMS-I
I Can't Go For That (No Can Do)	Daryll Hall & John Oates	–56 dB Peak –60.2 dB RMS-I
Rocket Man (I Think It's Going To Be A Long, Long Time)	Elton John	–59 dB Peak –64.4 dB RMS-I
Stayin' Alive	Bee Gees	–62.9 dB Peak –63.8 dB RMS-I

NOTE:

Full duration programme material used.

Processed using REAPER and JS: Mid/Side Encoder.

***diff* MS** = $\mathrm{MS}_{\mathrm{ouput}}$**Left** − $\mathrm{MS}_{\mathrm{ouput}}$**Right**

***diff* PAN** = $\mathrm{PAN}_{\mathrm{ouput}}$**Left** − $\mathrm{PAN}_{\mathrm{ouput}}$**Right**

7.2 Supplementary Figures

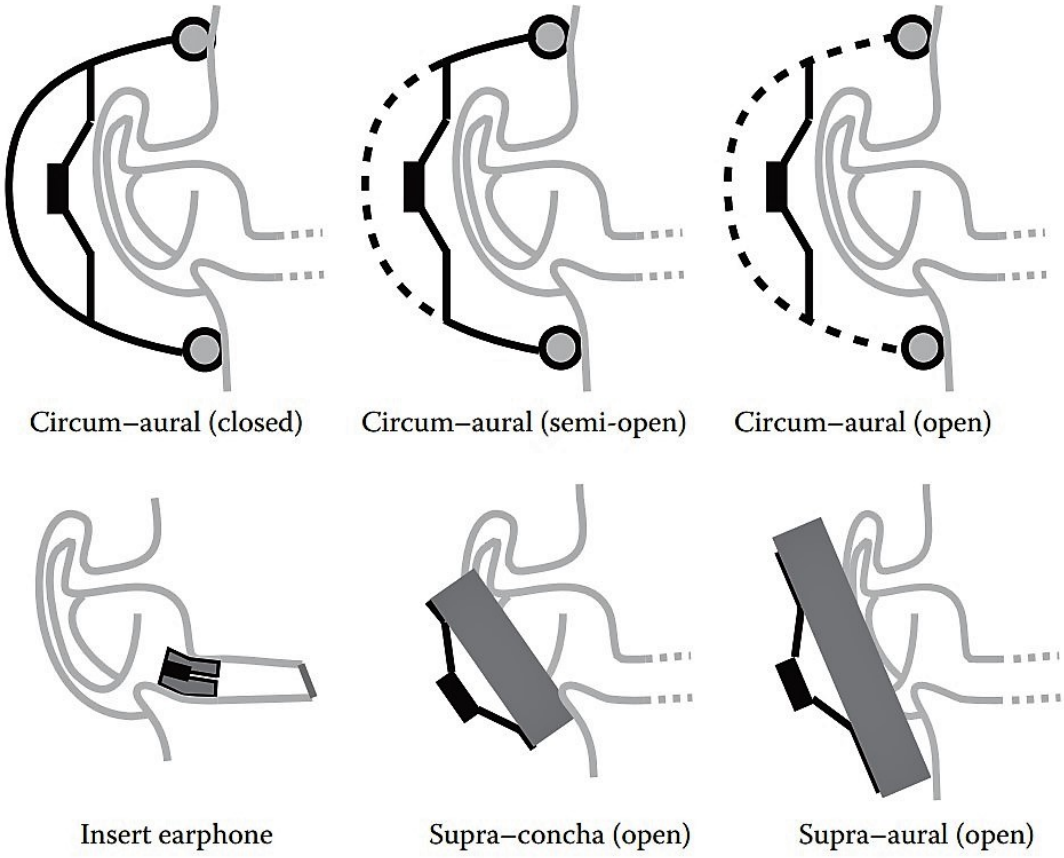

Figure 7.3: Diagrammatic Horizontal Sections Showing Types of Earphones. Reproduced from Kleiner M, 2013.[254(fig24.1)]

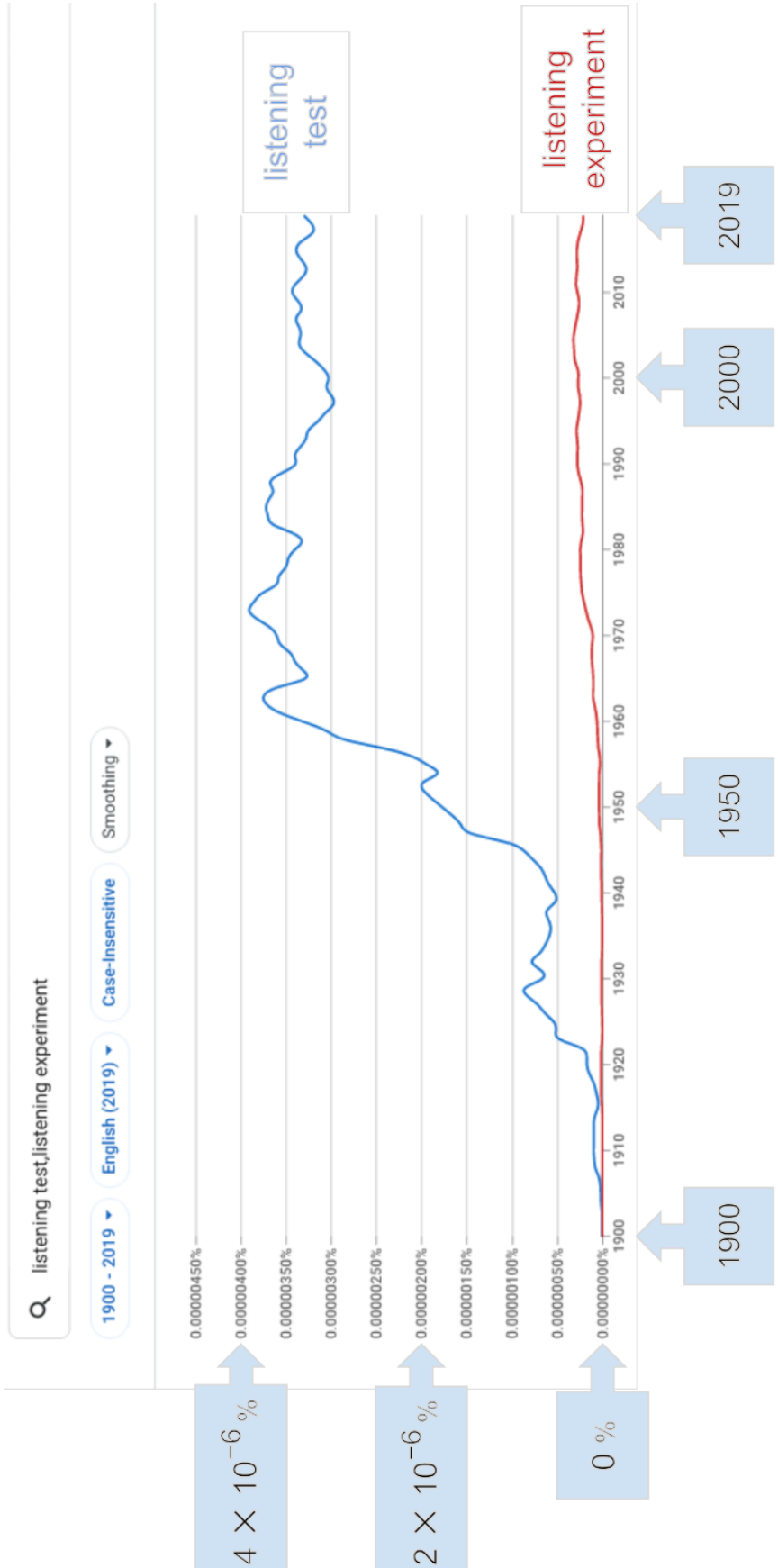

Figure 7.4: A graph showing how the phrases 'listening test' and 'listening experiment' have occurred in the *English 2019* corpus of books from 1900 to 2019 (*Google Books Ngram Viewer: 'listening test', 'listening experiment'*, 2021).[154,255]

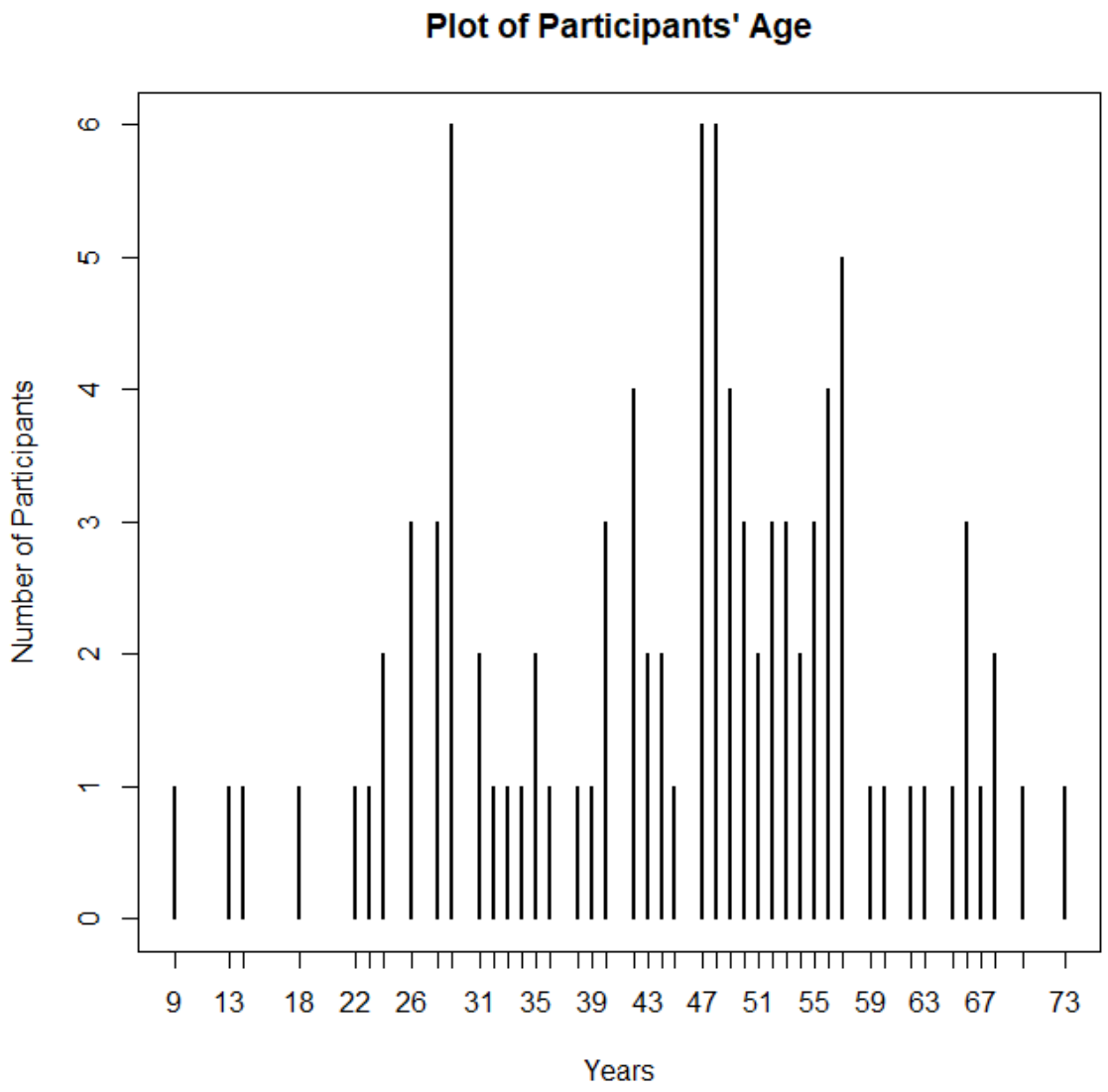

Figure 7.5: Plot showing Participants' Age Distribution (Years)

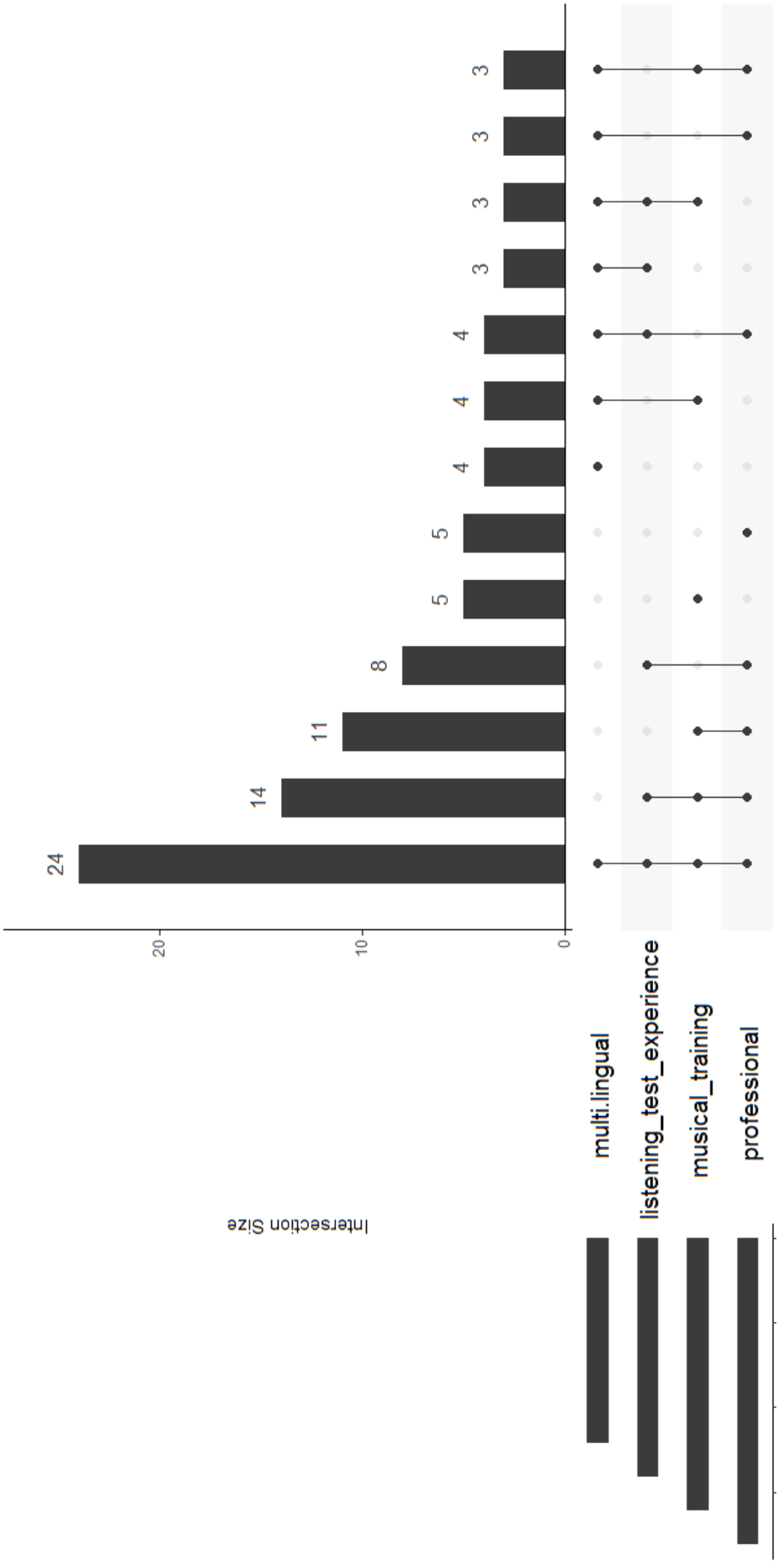

Figure 7.6: *UpSet*[225] Graph #1, showing Relationships Between Subject Groups

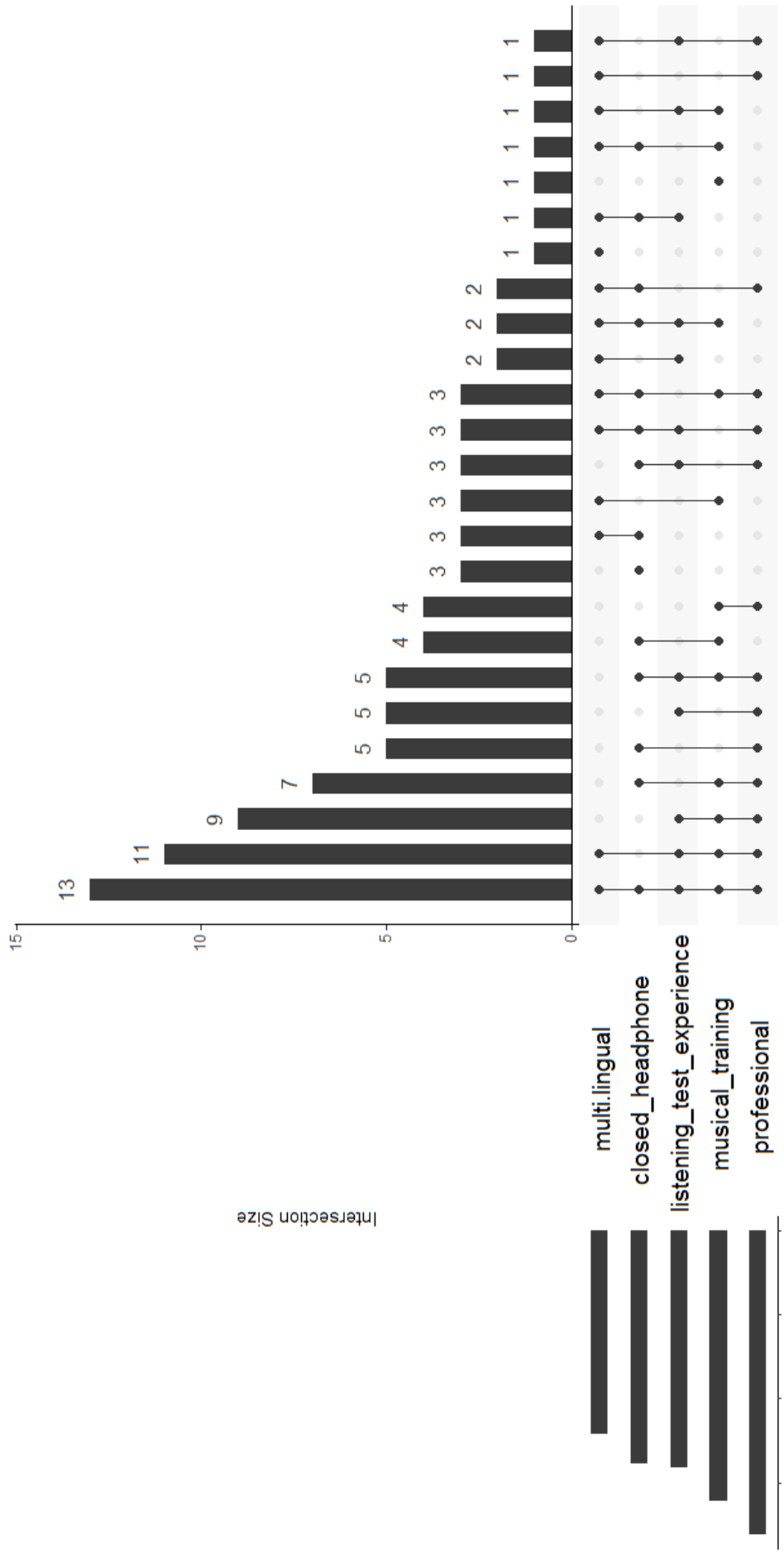

Figure 7.7: *UpSet*[225] graph #2, showing Relationships Between Subject Groups

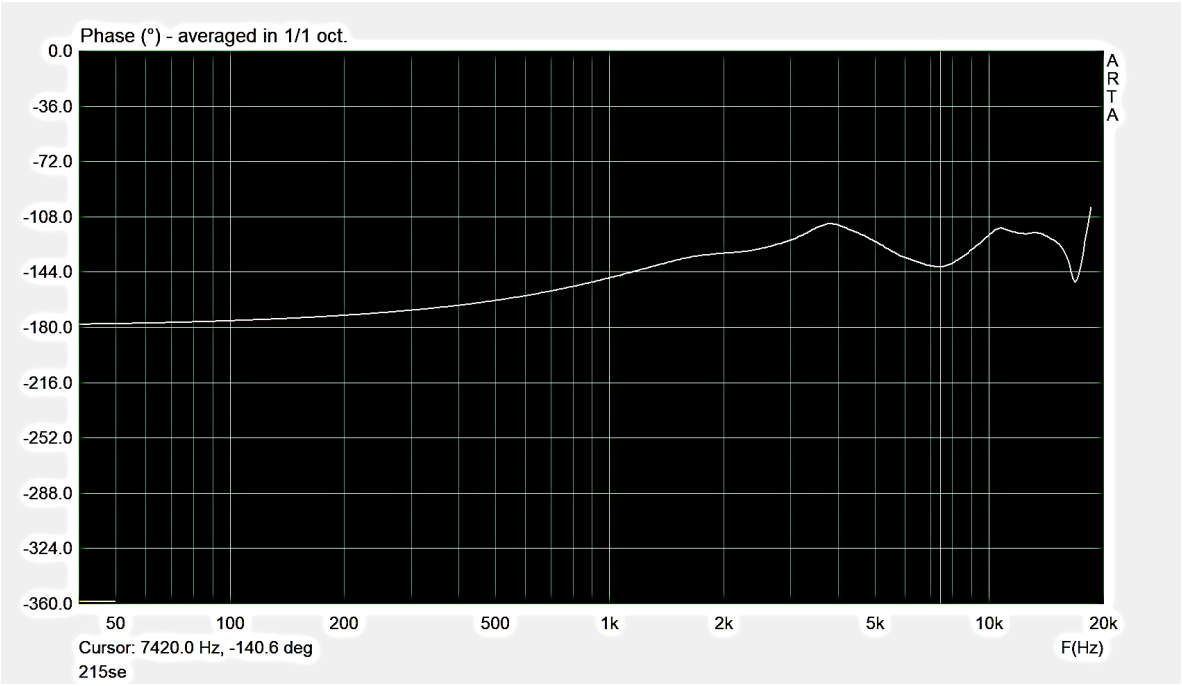

Figure 7.8: Earphone #1 phase estimate (note `Phase` range)

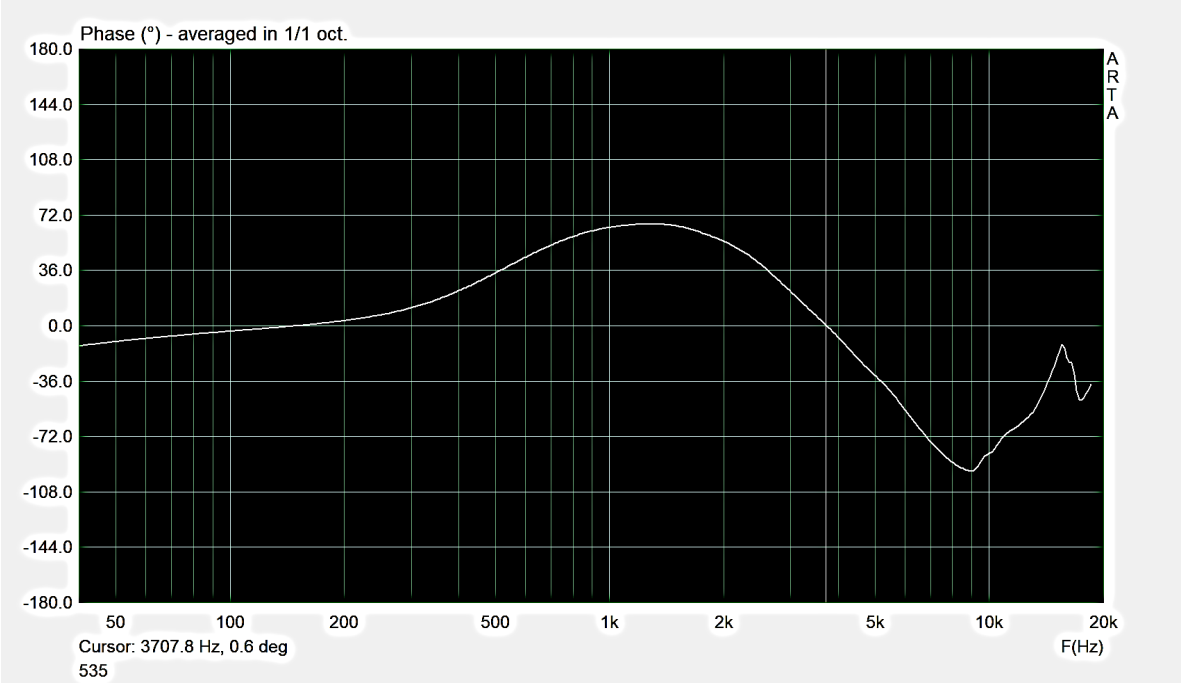

Figure 7.9: Earphone #2 phase estimate

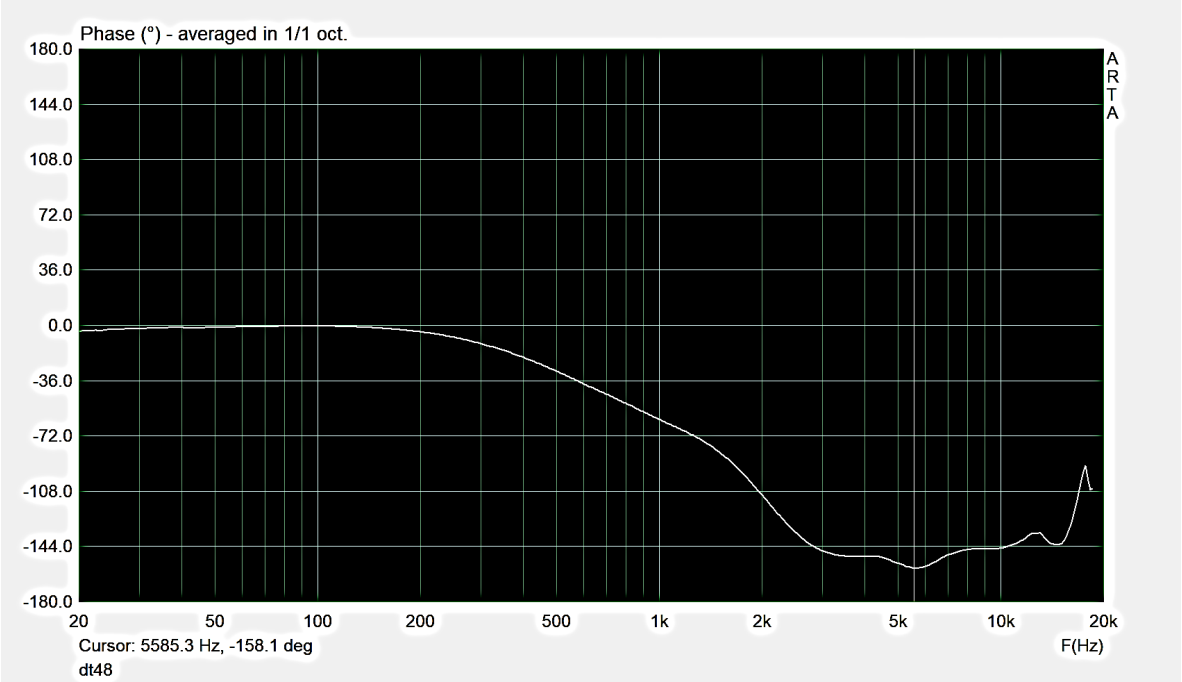

Figure 7.10: Earphone #3 phase estimate

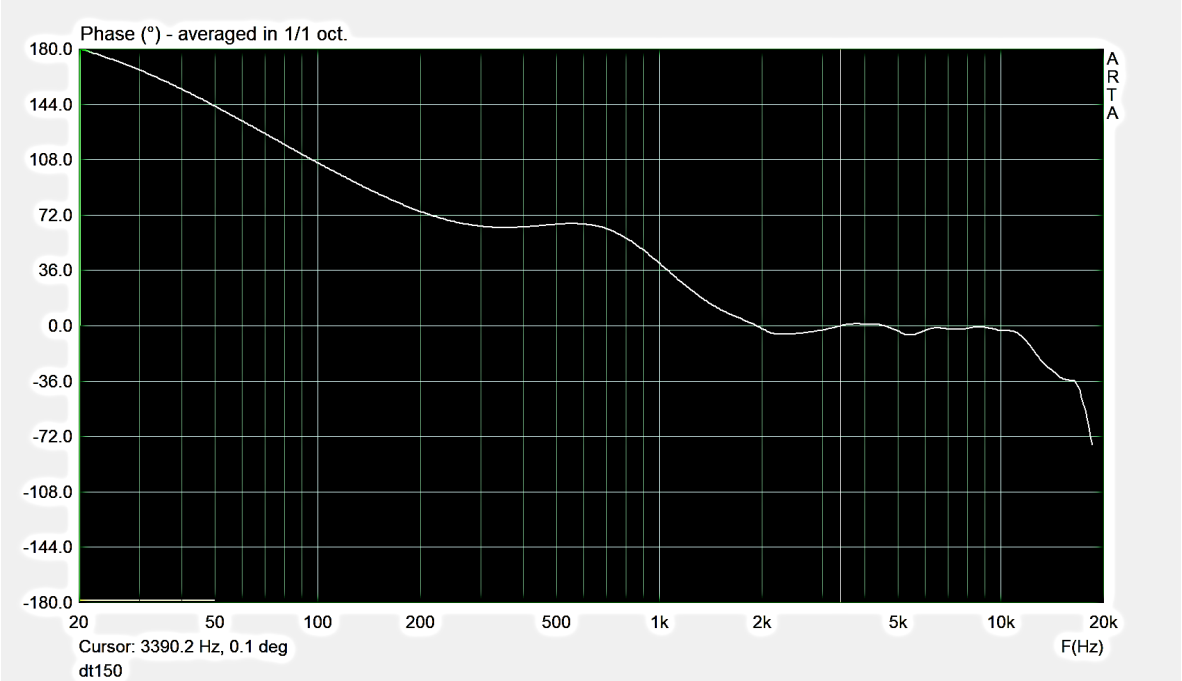

Figure 7.11: Earphone #4 phase estimate

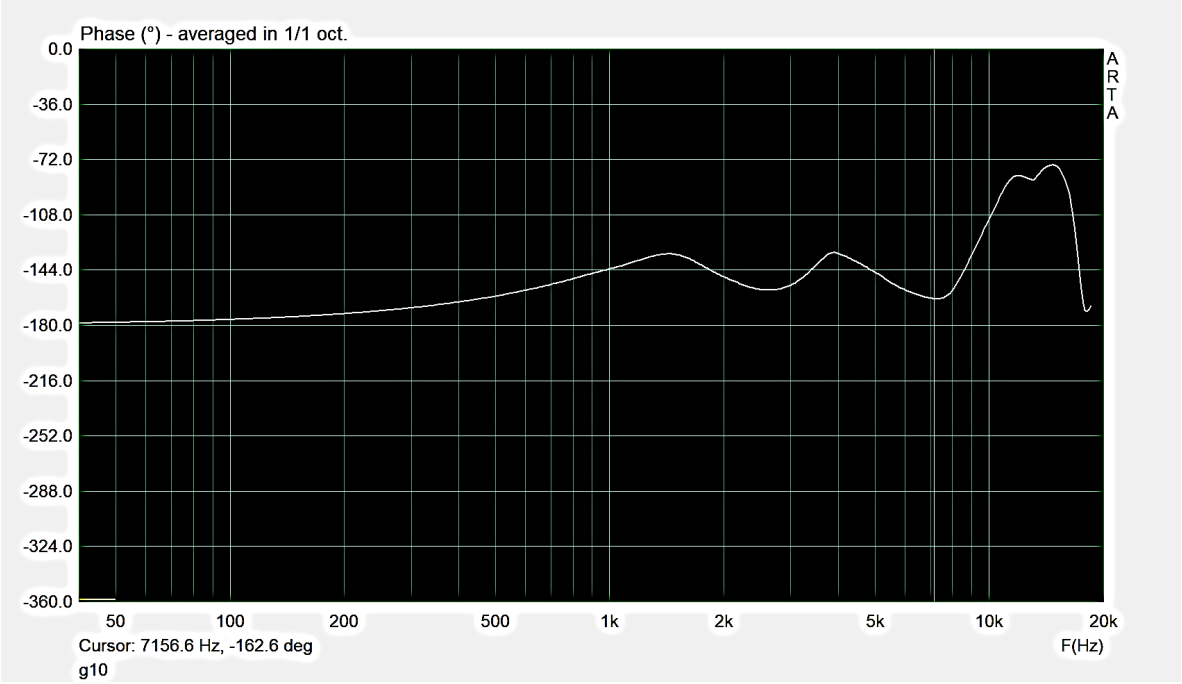

Figure 7.12: Earphone #5 phase estimate (note `Phase` range)

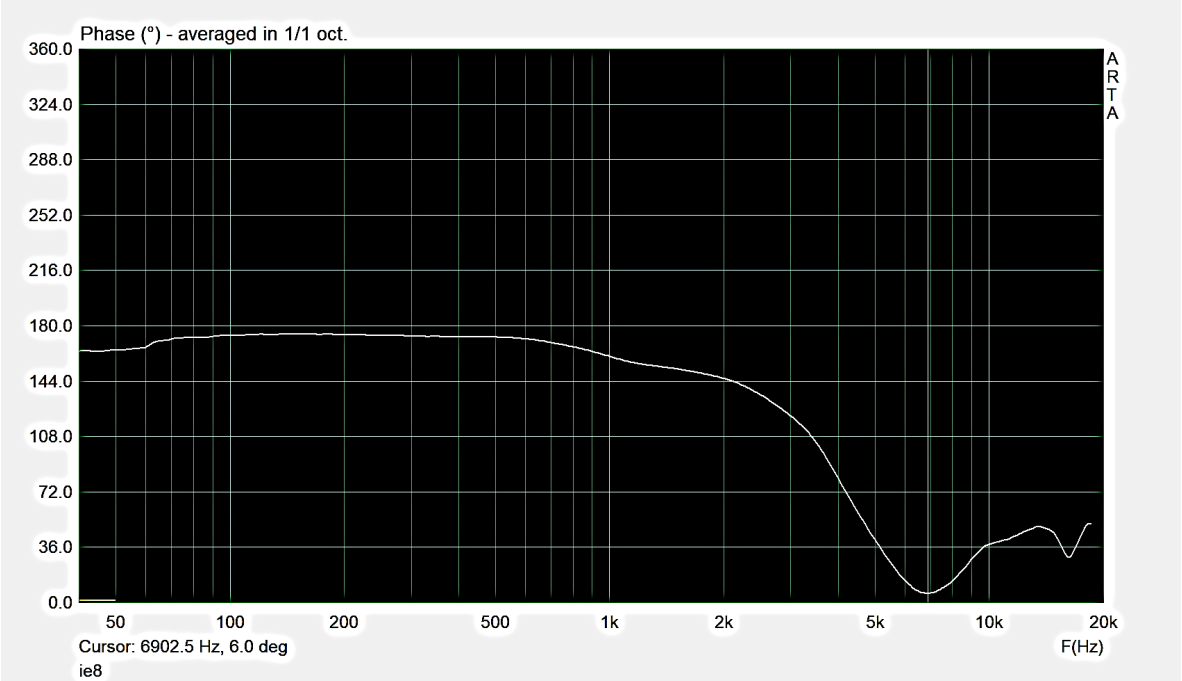

Figure 7.13: Earphone #6 phase estimate (note `Phase` range)

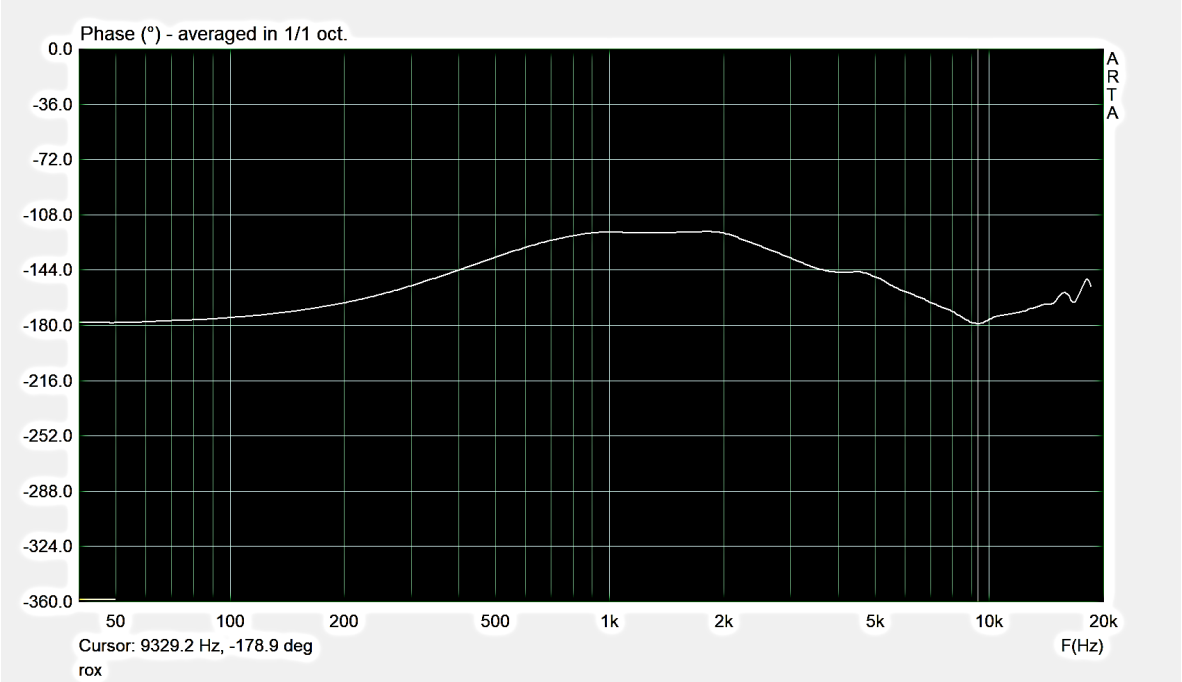

Figure 7.14: Earphone #7 phase estimate (note `Phase` range)

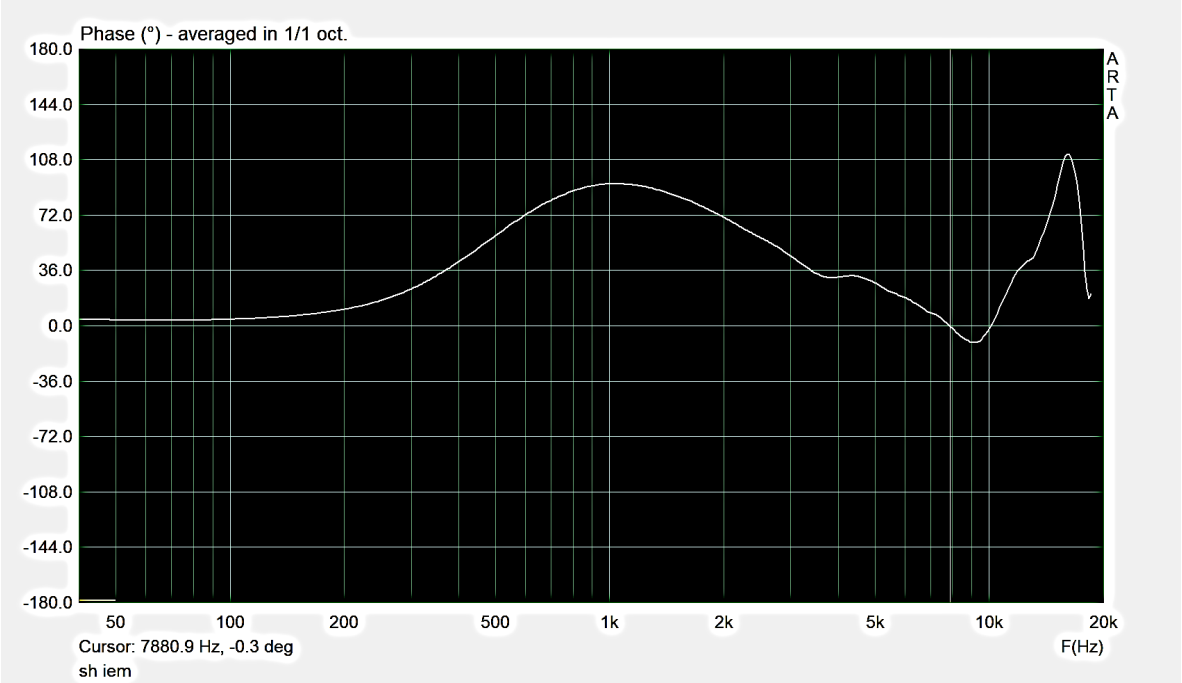

Figure 7.15: Earphone #8 phase estimate

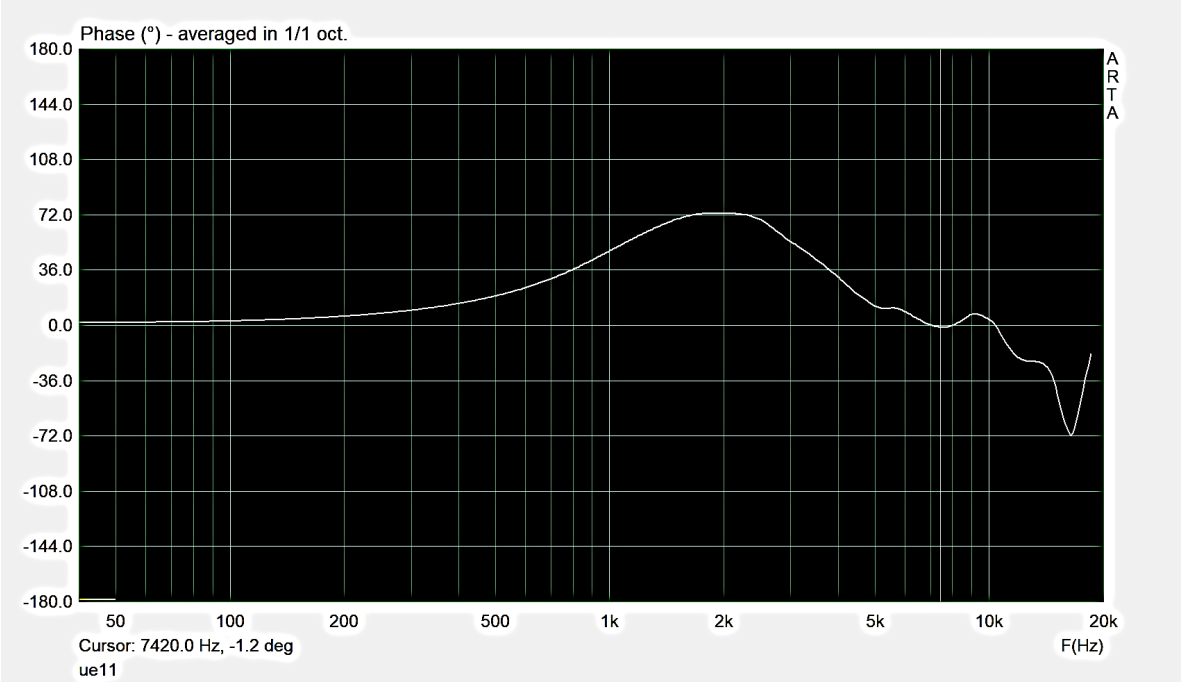

Figure 7.16: Earphone #9 phase estimate

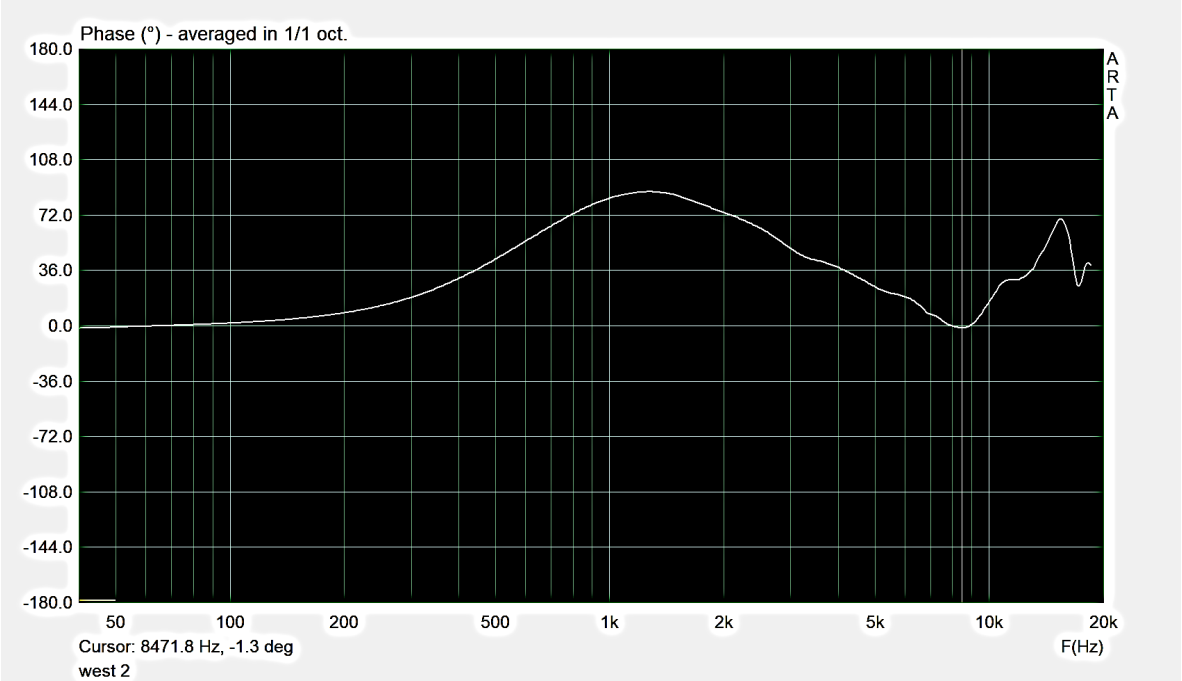

Figure 7.17: Earphone #10 phase estimate

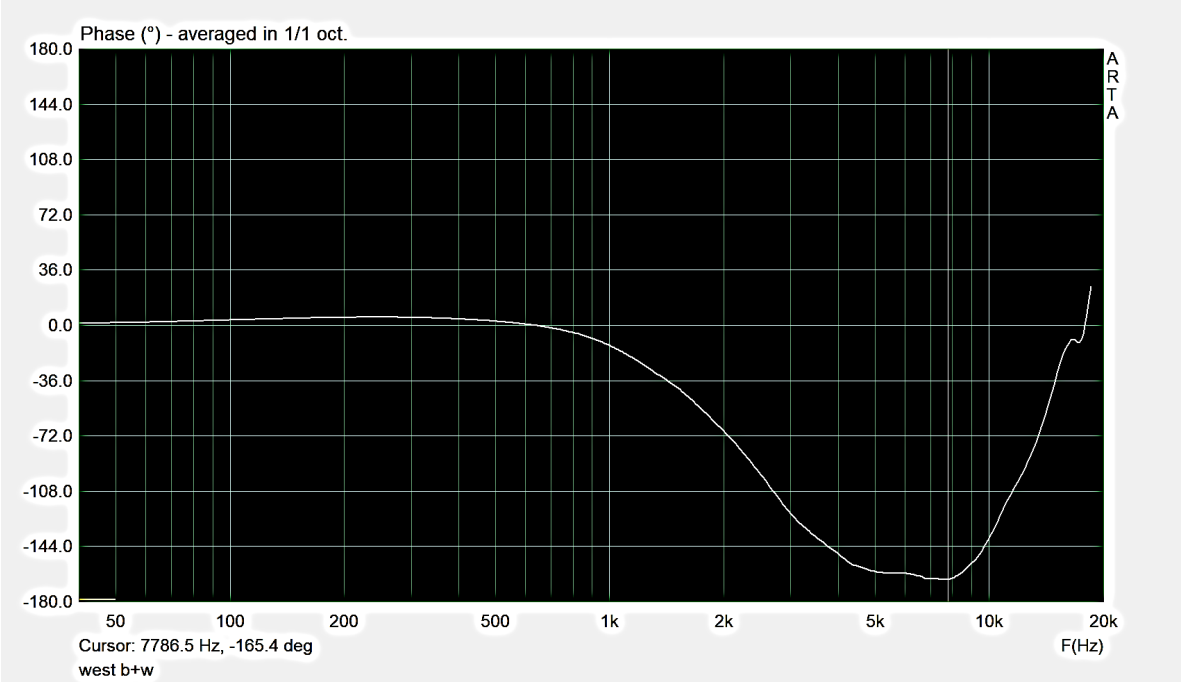

Figure 7.18: Earphone #11 phase estimate

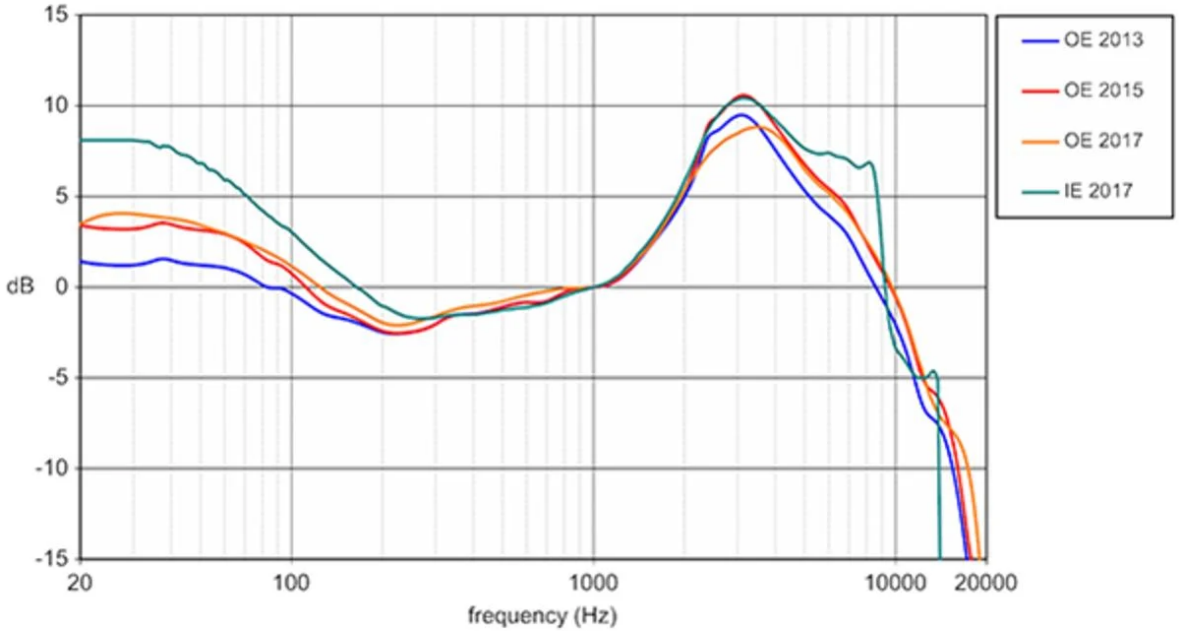

Figure 7.19: This graph compares the Harman curves published over the years, showing substantial variation in the bass and treble. Notably, the in-ear target [IE] reflects the difference in a blocked ear canal makes, as the blocked canal has it[s] own sets of resonances that the IEM needs to compensate for. (From: Harman). Note: OE — Over the Ear headphone; IE — In-Ear earphone. Reproduced from Ilkless, 2021.[256]

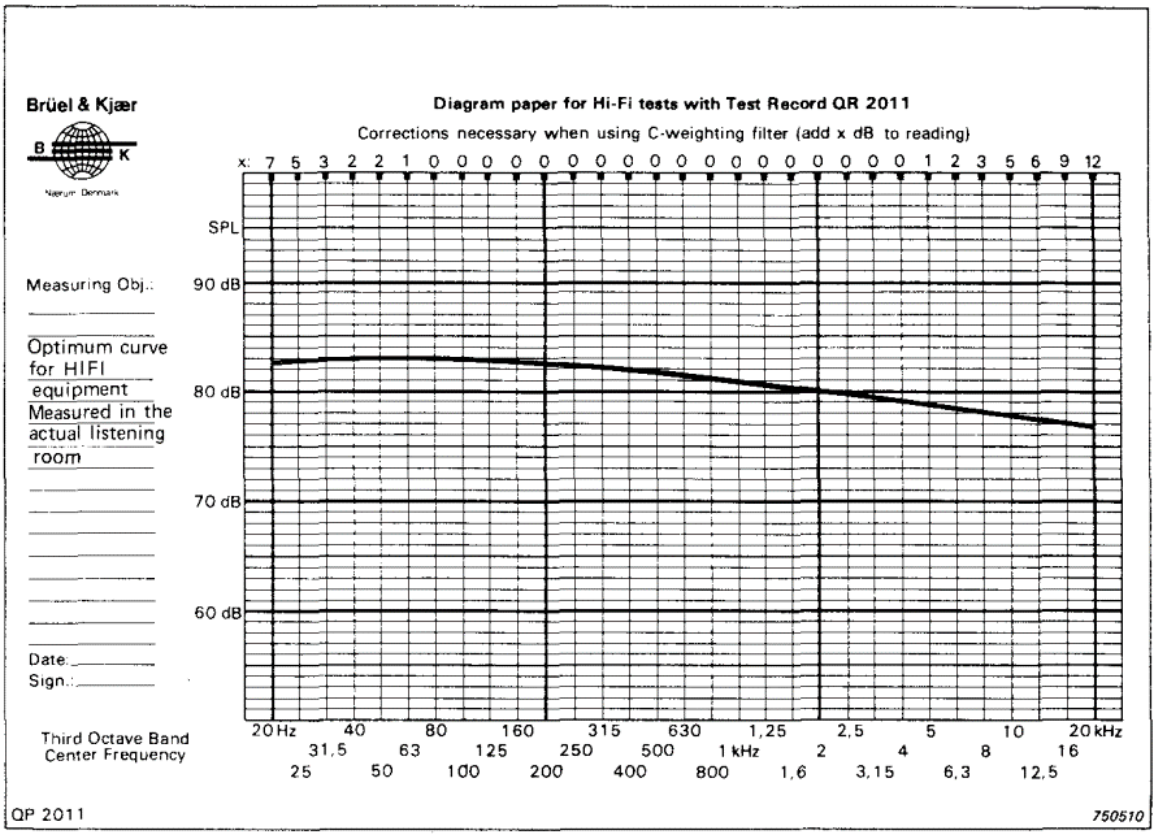

Figure 7.20: 'Optimum curve for HIFI equipment. Measured in the actual listening room'.[93(fig5)] Reproduced from Möller H, 1974.

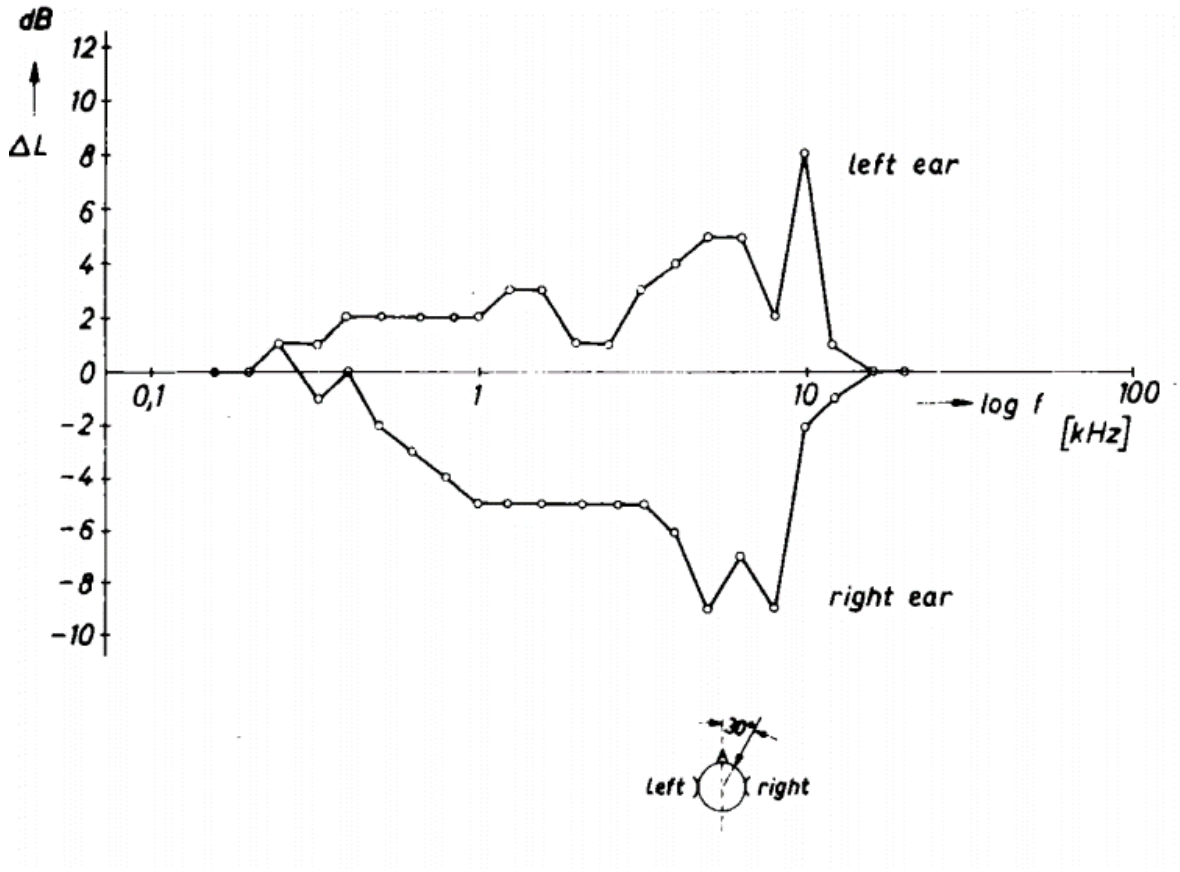

Figure 7.21: Simulation of the directional characteristic of the human head related to the sound incident direction (30°). Reproduced from Plenge G, 1974.[257(fig8)]

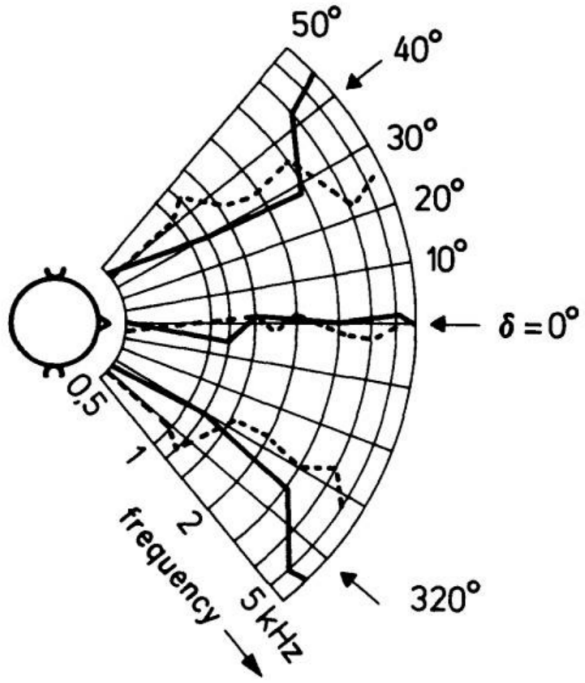

Figure 7.22: Direction of incidence of sound for sinusoids and Gaussian tone bursts of critical bandwidth so that the same direction of the auditory event is achieved as with broad band noise from the directions δ = 0', 40°, 320°. Solid lines: sinusoidal signals (after Sandel et al. 1955[a] sound pressure level 35 dB, 5 subjects, head immobilized). Dashed lines: Gaussian tone bursts (after Boerger 1965a[b],b[c];50-60 phon, 4–7 subjects, no head movement). Reproduced from Blauert J, 1996.[63(fig2.3)]

[a]Sandel, T. T., D. C. Teas, W. E. Feddersen, and L. A. Jeffress (1955): Localization of sound from single and paired sources. J. Acoust. Soc. Amer. 27, 842-852.

[b]Boerger, G. (1965a): Die Lokalisation von Gausstonen [The localization of Gaussian tones]. Dissertation, Technische Universitat, Berlin.

[c]Boerger, G. (1965b): Uber die Tragheit des Gehdrs bei der Richtungsempfindung [On the persistence of the directional sensation in hearing]. Proceedings, 5th Int. Congr. on Acoustics, Liege, B 27

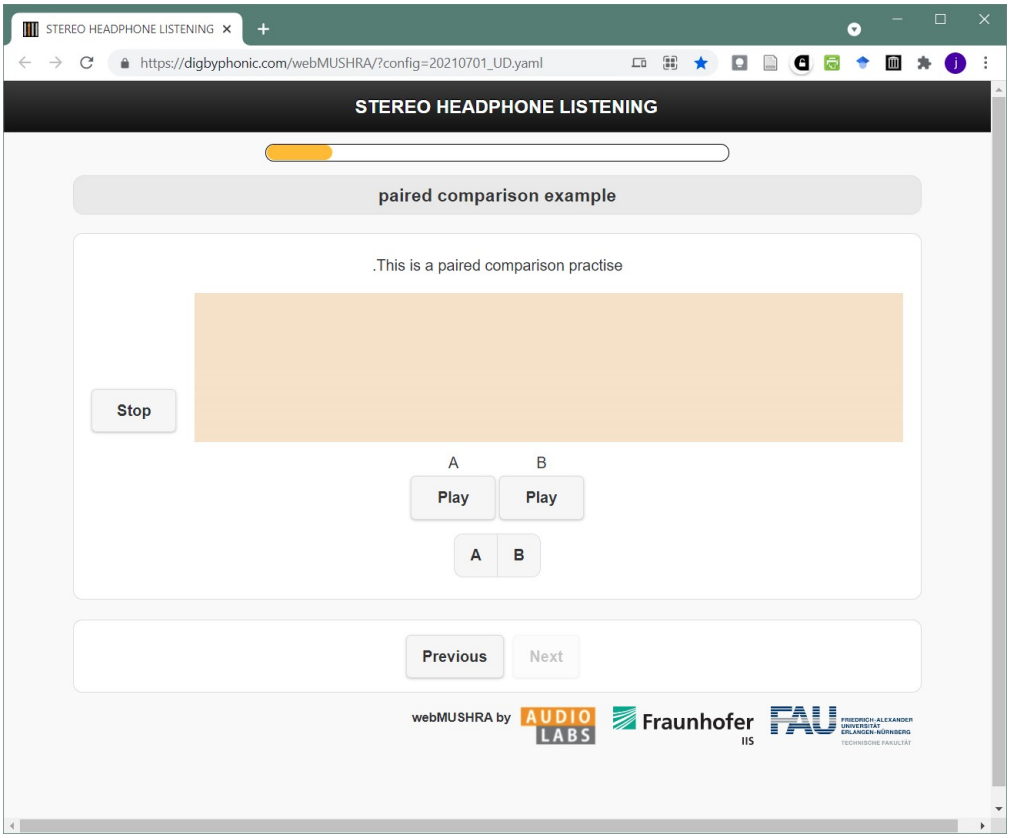

Figure 7.23: *webMUSHRA* 'paired comparison example' page, personal computer screenview.

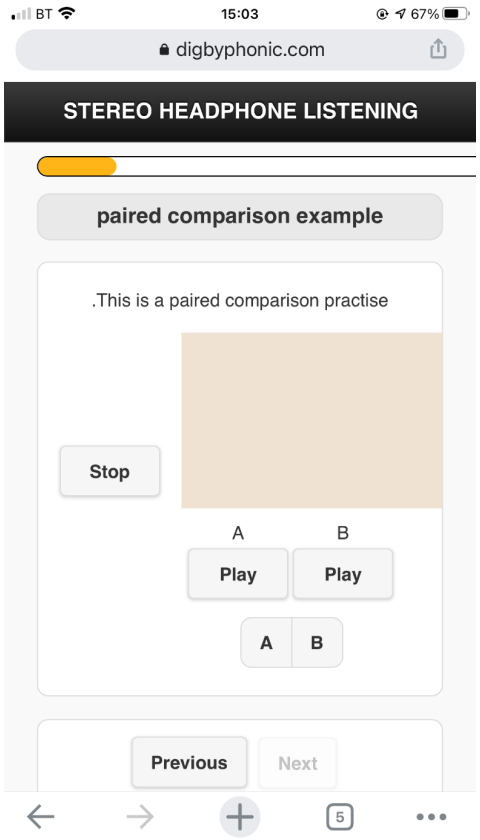

Figure 7.24: *webMUSHRA* 'paired comparison example' page, smartphone screenview.

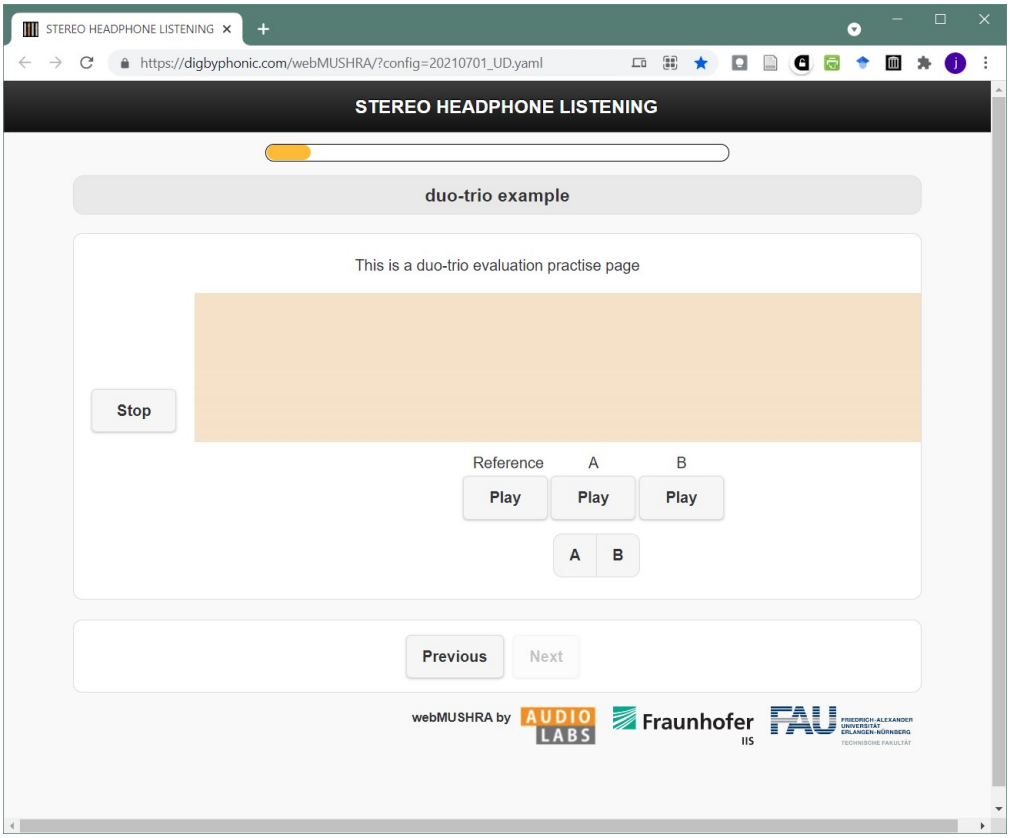

Figure 7.25: *webMUSHRA* 'duo-trio example' page,personal computer screenview.

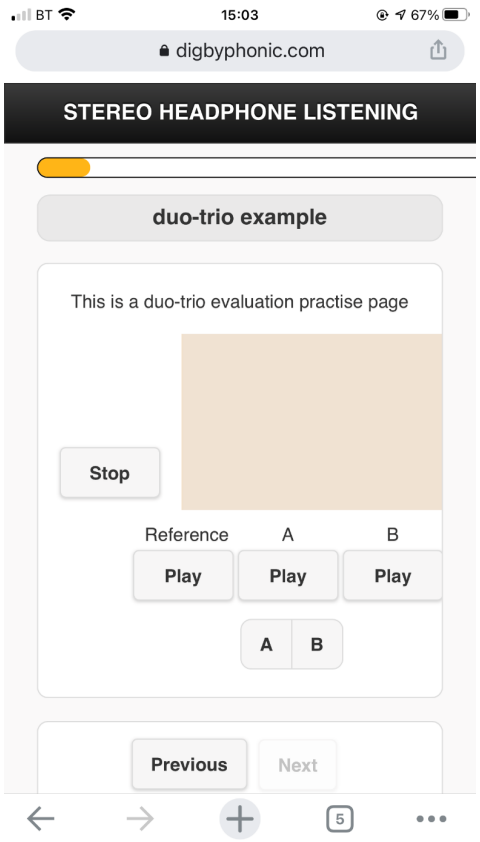

Figure 7.26: *webMUSHRA* 'duo-trio example' page, smartphone screenview.

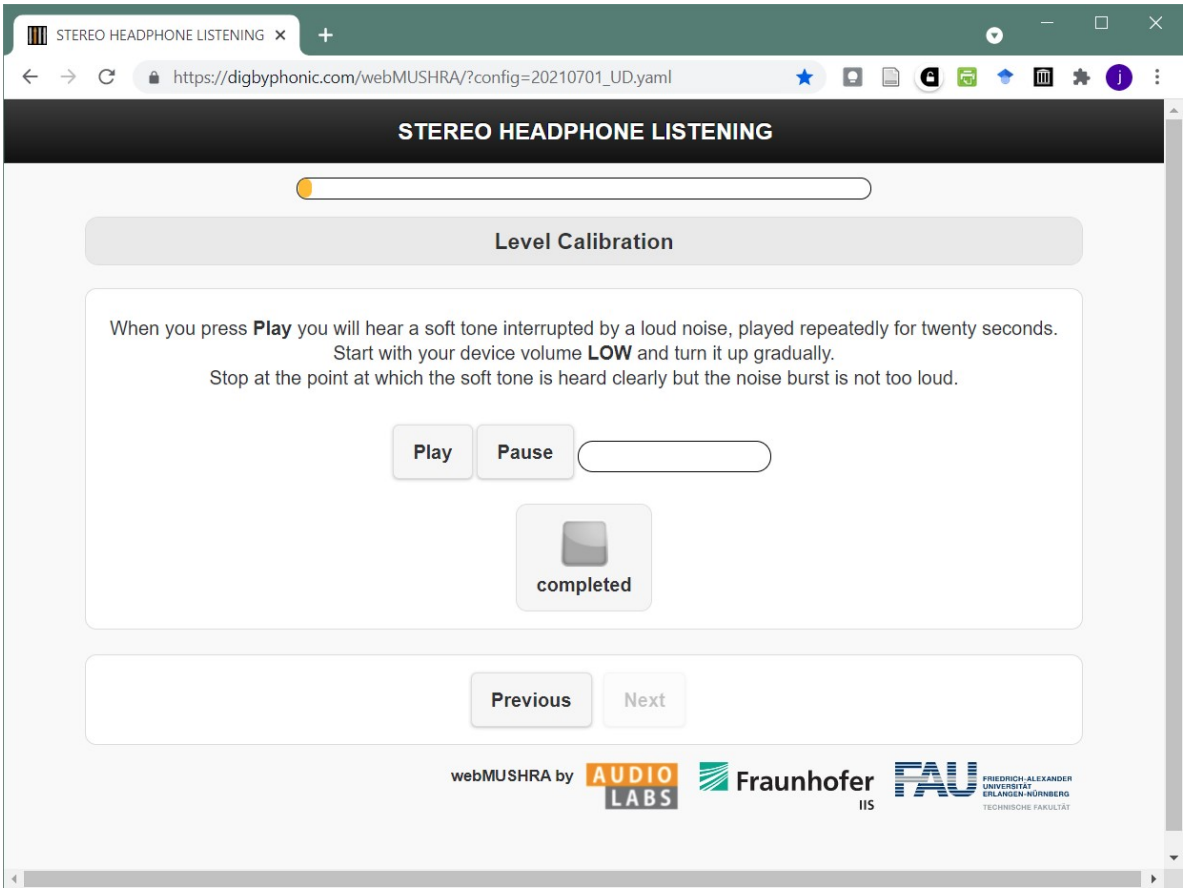

Figure 7.27: *webMUSHRA* 'Level Calibration' page, personal computer screenview

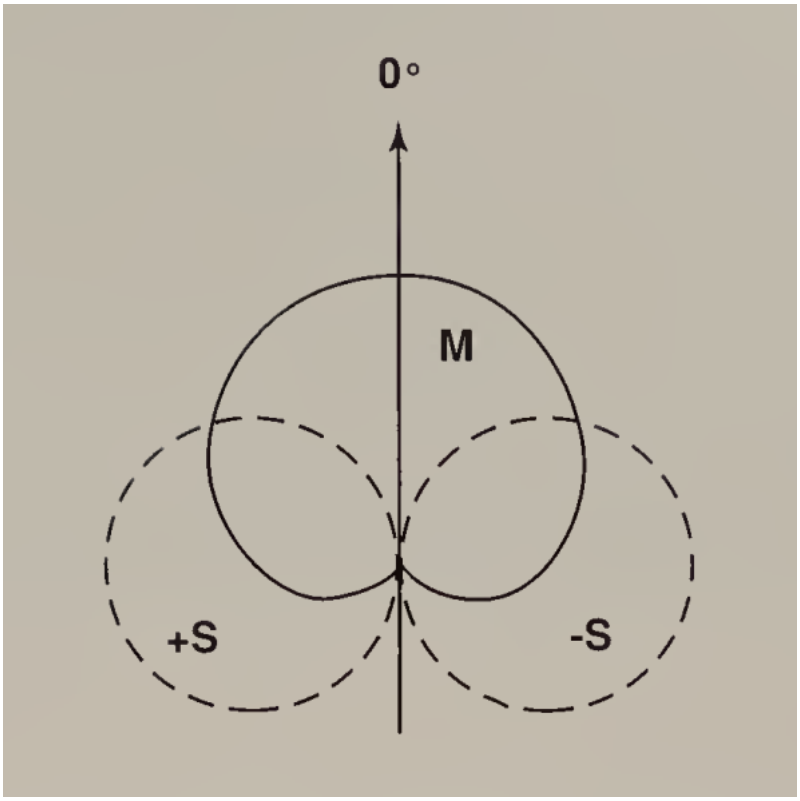

Figure 7.28: The basic configuration of a Mid/Side stereo pickup; the mid microphone is aimed at the 0° axis. The side microphone is a bidirectional, with its null plane aligned with the principal axis of the mid microphone. By convention, the positive lobe of the side microphone is oriented to the left side of the stereo soundstage. Reproduced from Streicher R, Everest FA; 1998.[73(fig7.17)]

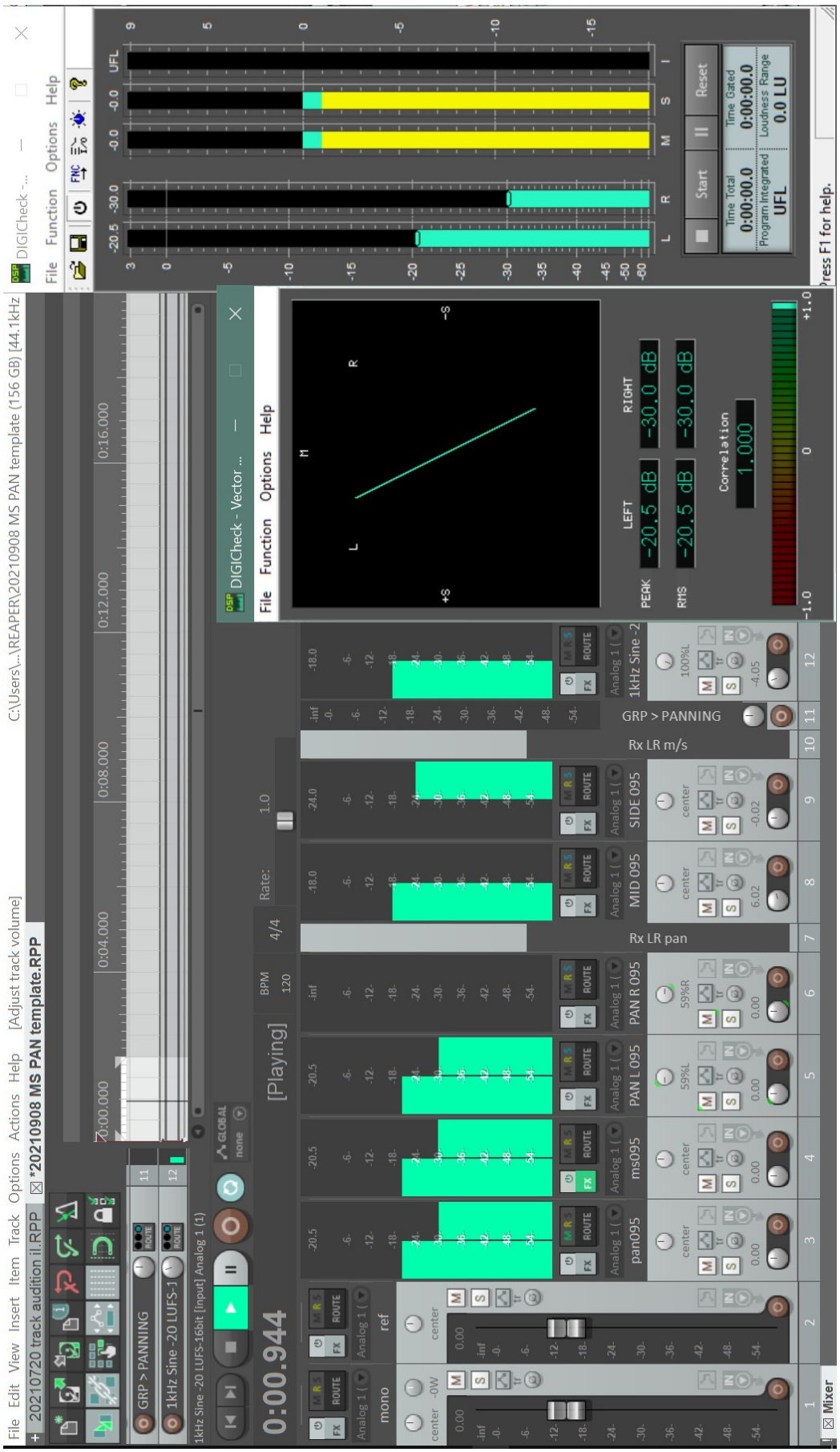

Figure 7.29: Example REAPER mix view showing relative crosstalk level for 1 kHz Sine routed to Left channel only of MS and PAN group busses.[172] Metered at PAN bus output using RME Digicheck.[202]

7.3 Experiment Invitation – Email

{EMAIL SUBJECT}

Stereo Headphone Listening Experiment

{DATE}

hello,

i would appreciate it if you could please take part in a headphone listening experiment for my University dissertation.

you'll need a pair of headphones or earphones connected to a device running a web browser. (laptop, computer, smartphone, tablet, etc.)

it will usually take 15 - 20 minutes total.

six participants will be chosen at random to receive a £25 gift voucher.

the reported data will only ever be processed as a group, never individually, and will always remain strictly anonymous and confidential.

if you decide to withdraw your results from the experiment simply email me before JULY 31st, 2021.

it is not a test, listening preferences are all that\'s required, there are no wrong answers.

click the link to participate, allowing time for the audio to load, indicated by the rotating disk :
https://digbyphonic.com/webMUSHRA/?config=20210701_UD.yaml

feel free to forward this invite to anybody who may be interested, it would be appreciated by all, and will help toward a successful outcome.

thank you kindly

best wishes

jonathan digby

University of Derby

j.digby1@unimail.derby.ac.uk

7.4 Determining Number of Assessors

Reproduced from Annex section 'A3' of: International Organization for Standardization. ISO 10399:2018 Sensory analysis – Methodology – Duo- trio test. CEN; 2018.

Table in Figure 7.30 shows a statistical approach for determining the number of assessors. The statistical sensitivity of the test is a function of three values: the α-risk, the β-risk, and the maximum allowable proportion of distinguishers, p_d. Prior to conducting the test, select values for α, β and p_d using the following guidelines.

NOTE In this document, the probability of a correct response, p_c, is modelled as $p_c = p_c + (1 - p_d)/2$, where p_d is the proportion of the population of assessors who can distinguish between the two products. A psychometrical model of the assessor's decision process, such as the Thurstone-Ura model, could also be applied in a duo-trio test.

As a rule of thumb, a statistically significant result at

— an α-risk of 10 % to 5 % (0,10 to 0,05) indicates slight evidence that a difference was apparent,

— an α-risk of 5 % to 1 % (0,05 to 0,01) indicates moderate evidence that a difference was apparent,

— an α-risk of 1 % to 0,1 % (0,01 to 0,001) indicates strong evidence that a difference was apparent, and

— an α-risk below 0,1 % ($< 0{,}001$) indicates very strong evidence that a difference was apparent.

For β-risks, the strength of the evidence that a difference was not apparent is assessed using the same criteria as above (substituting "β" for "α" and "was not apparent" for "was apparent"). The maximum allowable proportion of distinguishers, p_d, falls into three ranges:

— $p_d < 25\%$ represents small values;

— $25\% \leq p_d \leq 35\%$ represents medium-sized values;

— $p_d > 35\%$ represents large values.

Choose the number of assessors so as to obtain the level of sensitivity required for the test. Enter table in Figure 7.30 in the section corresponding to the selected values of pd and the column corresponding to the selected value of β. The minimum required number of assessors is found in the row corresponding to the selected value of α. Alternatively, table in Figure 7.30 may be used to develop a set of values for p_d, α and β that provide acceptable sensitivity while maintaining the number of assessors within practical limits. The approach is presented in detail in:

Schlich P. Risk tables for discrimination tests. Food Quality and Preference. 1993;4(3):141–151.

Values given in the table in Figure 7.30 are the minimum number of assessors required to execute a duo-trio test with a specified sensitivity determined by the values of p_d, α and β. Enter the table in the section corresponding to the chosen value of p_d and the column corresponding to the chosen value of β. Read the minimum number of assessors from the row corresponding to the chosen value of α.

α	p_d	β				
		0,20	**0,10**	**0,05**	**0,01**	**0,001**
0,20	50 %	12	19	26	39	58
0,10		19	26	33	48	70
0,05		23	33	42	58	82
0,01		40	50	59	80	107
0,001		61	71	83	107	140
0,20	40 %	19	30	39	60	94
0,10		28	39	53	79	113
0,05		37	53	67	93	132
0,01		64	80	96	130	174
0,001		95	117	135	176	228
0,20	30 %	32	49	68	110	166
0,10		53	72	96	145	208
0,05		69	93	119	173	243
0,01		112	143	174	235	319
0,001		172	210	246	318	412
0,20	20 %	77	112	158	253	384
0,10		115	168	214	322	471
0,05		158	213	268	392	554
0,01		252	325	391	535	726
0,001		386	479	556	731	944
0,20	10 %	294	451	618	1 006	1 555
0,10		461	658	861	1 310	1 905
0,05		620	866	1 092	1 583	2 237
0,01		1 007	1 301	1 582	2 170	2 927
0,001		1 551	1 908	2 248	2 937	3 812

Figure 7.30: Number of assessors needed for a duo-trio test. Adapted from Schlich P, 1993.[226]

7.5 The *R* Project for Statistical Computing

7.5.1 *R* Code – Examples

Load Wilcox custom function package 'Rallfun-v28', import CSV file from *webMUSHRA* into an *R* data frame, check import:

```
> source("Rallfun-v28") # Wilcox R package custom functions
```

```
> dat = read.csv(file.choose(), header = TRUE, stringsAsFactors =
TRUE, check.names = TRUE) # create data frame from CSV
```

```
> str(dat) # data frame columns
```

which returns

```
'data.frame': 3264 obs. of 17 variables:
$ session_test_id : Factor w/ 1 level
$ sex : int
$ volume_level : int
$ headphone_type : int
$ musical_training : int
$ multi.lingual : intbb
$ listening_test_experience: int
$ professional : int
$ feedback : Factor w/ 56 levels
$ age : Factor w/ 47 levels
$ email : Factor w/ 96 levels
$ trial_id : Factor w/ 34 levels
$ choice_reference : Factor w/ 1 level
$ choice_non_reference : Factor w/ 1 level
$ choice_answer : Factor w/ 2 levels
$ choice_time : int
$ choice_comment : logi
```

Remove text from 'age' variable vectors, convert to numeric, round down:

```
> dat\$age = as.numeric(gsub(\".\*?(\[0-9\]+).\*\", \"\\\\1\",
dat\$age))
```

```
> summary(dat\$age)
```

which returns the Minimum, 1st Quartile, Median, Mean, 3rd Quartile, and Maximum

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
9.00	33.75	47.50	44.81	55.00	73.00

Create UpSet figure in report:

```
> flag = dat\$trial_id == \'12ms1_paired\' # variable single trial
```

```
> upset(dat\[flag,\],sets = c(\"musical_training\", \"multi.lingual\",
\"listening_test_experience\", \"professional\"), order.by =
\'freq\', text.scale = c(1.3,1.3,1,1,2,2) # outputs UpSet figure
```

Create histogram of four headphone choices:

```
> hist(dat\$phone\[flag\], labels = c(\"Closed-back\",\"Open-back\",\"In-Ear\",
ylim = c(0,60), breaks = c(0,1,2,3,4), main = \"Histogram of
Headphone Type\", xlab = \"\", ylab = \"Distribution\")
```

Create global variable for group answering yes to the four questions:

```
> z = dat\$multi.lingual + dat\$listening_test_experience +
dat\$musical_training + dat\$professional # sum responses
```

```
> id_insect = z > 3    # global variable for group answering yes to
the four answers
```

Create vector from text answers, ‘correct’ = 1, ‘incorrect’ = 0:

```
> dat\$ans = ifelse(dat\$choice_answer == \"correct\", 1, 0)
```

Select four trial instances starting with the prefix “12ms”:

```
> x = grepl(\"12ms\", dat\$trial_id)
```

```
> head(dat\$trial[x]) # shows first parts of vector
```

which returns the first six vectors, confirming all four replication versions as selected

```
[1] 12ms2_paired 12ms1_paired 12ms3_paired 12ms4_paired 12ms1_paired
```

```
[6] 12ms4_paired
```

```
34 Levels: 12ms1_paired 12ms2_paired 12ms3_paired 12ms4_paired ...
pol8r2_duotrio
```

Calculate one-sample case for x using binom.conf function:

```
> dat\$trial_id\[x\] # verifies the 384 chosen vectors. (not shown)

> binom.conf(sum(dat\$ans\[x\]),length(dat\$ans\[x\]), alpha =
0.05)
```

which returns: $\widehat{p}$ (estimate of p); estimated standard error; confidence interval; n observations

```
$phat

[1] 0.4765625

$se

[1] 0.02536148

$ci

[1] 0.4270870 0.5265022

$n

[1] 384
```

Calculate one-sample case for 'intersect' group for x:

```
> x_insect = x & id_insect # logical query

> binom.conf(sum(dat\$ans\[x_insect\]), length(dat\$ans\[x_insect\]),
alpha = 0.05) # calculate
```

Calculate one-sample case for null 'intersect' group:

```
> x_notinsect = x & !id_insect

> binom.conf(sum(dat\$ans\[x_notinsect\]),length(dat\$ans\[x_notinsect\]),
alpha = 0.05) # calculate
```

Calculate two-sample case with 2000 iterations:

```
> binom2g(sum(dat\$ans\[x_insect\]), length(dat\$ans\[x_insect\]),
sum(dat\$ans\[x_notinsect\]), length(dat\$ans\[x_notinsect\]),
alpha = 0.05, iter = 2000)
```

Plot of participants' ages:

```
> plot(table(dat\$age\[flag\]), ylab = \"Number of Participants\",
main = \"Plot of Participants\' Age\", xlab = \"Years\")
```

7.5.2 *R* code: Gift Voucher Winners

Pick gift voucher winners:

```
> names=unique(data\[,11\]) # vector containing unique entries from
col. 11, 'email'

> winners=sample(names,6) # 6 random selections

> winners # output values ('email')
```

which returns a list of six unique random values from column 11 of the dataframe [email].

Mersenne-Twister randomisation is used by default:

Matsumoto M. Mersenne Twister: A random number generator (since 1997/10). 2021 Aug 1 [accessed 2026 Oct 24]. http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/emt.html

R uses its own initialization method due to B. D. Ripley and is not affected by the initialization issue in the 1998 code of Matsumoto and Nishimura addressed in a 2002 update:

R Core Team. R: A language and environment for statistical computing. 2025 [accessed 2026 Oct 24]. https://www.R-project.org/

7.5.3 Agresti-Coull Method

Wilcox RR. Understanding and applying basic statistical methods using R. John Wiley & Sons; 2017. p.169

The explanation of the Agresti-Coull method as applied in the Wilcox *R* functions acbinomci and `binom.conf` is reproduced from page 169 in the source:

Let x represent the observed number of successes among n observations, in which case

$$\widehat{p} = \frac{x}{n}$$

is the proportion of successes among the n observations. Let c be the $1 - \alpha/2$ quantile of a standard normal distribution. Compute

$$\tilde{n} = n + c^2$$

$$\tilde{x} = x + \frac{c^2}{2}$$

and

$$\tilde{p} = \frac{\tilde{x}}{\tilde{n}}$$

Then the $1 - \alpha$ confidence interval for p is

$$\tilde{p} \pm c\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

7.5.4 Computational Details

The statistical results in this paper were obtained using *R* version 4.1.0 (2021-05-18) with the packages `Rallfun-v38`, `akima 0.6-2.1`, `MASS 7.3-54`, `mgcv 1.8-36`, `plotrix 3.8-1`, `quantreg 5.86`, `robust 0.5-0.0`, `rrcov 1.5-5`, `scatterplot3d 0.3-41`, `stats 4.1.0`, `UpSetR 1.4.0`.

R itself and all packages used are available from CRAN at https://cran.r-project.org/

R package `Rallfun-v38` includes functions `binom.conf` and `acbinomci` used for this report:

Wilcox RR. Software > Rand Wilcox > USC Dana and David Dornsife College of Letters, Arts and Sciences. 2021 Mar [accessed 2026 Oct 24]. https://osf.io/xhe8u/

7.6 Auditory Demonstration and Test Signals – Sources

30th Anniversary Sampler [CD] (2006). USA: Reference Recordings (RR-908).

Best of Chesky Classics & Jazz And Audiophile Test Disc Volume 3 [CD] (1994). USA: Chesky Records (JD 111).

Best Of Chesky Jazz And More Audiophile Tests Volume 2 [CD] (1992). USA: Chesky Records (JD68).

Bregman, A.S. and Ahad, P.A. (1990) *Demonstrations to accompany Bregman's: Auditory Scene Analysis - The perceptual organization of sound* [CD]. USA: MIT Press.

CANFORD INTERNATIONAL BROADCAST STANDARDS TESTS CD [CD] (1998). UK: Canford Audio plc (25-321).

Chesky Records Jazz Sampler & Audiophile Test Compact Disc: Volume 1 [CD] (1990). USA: Chesky Records (Chesky JD37).

Deutsch, D. (1995) *Musical Illusions and Paradoxes* [CD]. USA: Philomel (Philomel-001).

Dr. Chesky's Stereophonic Recording Microphone Techniques [CD] (2019). USA: Chesky Records.

Dr. Chesky's Ultimate Headphone Demonstration [CD] (2014). USA: Chesky Records.

Euroradio Measurements Test Material [FLAC download] (2008). Geneva: EBU (Euroradio CD). Available at: https://tech.ebu.ch/publications/euroradio-CD (Accessed: 2026-10-22).

Houtsma, A.J.M. and Wagenaars, W.M. (1988) *Auditory Demonstrations* [CD]. UK, Europe & USA: Philips (1126-061).

Linkwitz, S. (2019a) *Phantom center*, *Linkwitz Lab*. Available at: https://www.linkwitzlab.com/Loudspeaker-Room/phantom_center.htm (Accessed: 2026-10-22).

Linkwitz, S. (2019b) *Room tests & measurements*, *Linkwitz Lab*. Available at: https://www.linkwitzlab.com/Loudspeaker-Room/tests&measurements.htm (Accessed: 2026-10-22).

Linkwitz, S. (2019c) *Test signal*, *Linkwitz Lab*. Available at: http://www.linkwitzlab.com/Ping_tests/Test_signal.htm (Accessed: 2026-10-22).

'My Disc' The Sheffield / A2TB Test Disc [CD] (1994). US: Sheffield Lab (The Audiophile Reference Series 10045-2-T).

NASA Glenn Research Center Acoustical Testing Laboratory (2003) *AUDITORY DEMONSTRATIONS II - Challenges in Speech Communication and Music Listening* [CD]. USA: NASA Glenn Research Center.

Nelson, D. and Taylor, A. (1997) *Acoustics, Audiometric Monitoring, Hearing Conservation, Noise Control* [CD]. USA: NASA Glenn Research Center.

Parameters For The Subjective Evaluation Of The Quality Of Sound – Music [FLAC download] (2008). Geneva: EBU (EBU PEQS CD). Available at: https://tech.ebu.ch/publications/PEQS_FLAC (Accessed: 2026-10-22).

Parsons, A. and Court, S. (1993) *Sound Check* [CD]. UK: Mobile Fidelity Sound Lab (APSC 2223).

Sound Quality Assessment Material Recordings For Subjective Tests [FLAC download] (2008). Geneva: EBU (EBU SQAM CD). Available at: https://tech.ebu.ch/publications/sqamcd (Accessed: 2026-09-02).

The Sheffield Lab Drum & Track Disc - For Audio Component Testing And Evaluation [CD] (2003). USA: Lim (The Legendary Sheffield Lab Series LIM XXR 005).

The Ultimate Demonstration: Chesky Records' Guide to Critical Listening [CD] (1995). USA: Chesky Records (UD95).

The Ultimate Demonstration Volume 2 [CD] (2008). USA: Chesky Records (SACD343).

7.7 Bregman's ' "Auditory Lake" Analogy' (2005)

The following passage is reproduced from:

Bregman AS. Auditory Scene Analysis and the Role of Phenomenology in Experimental Psychology. Canadian Psychology/Psychologie canadienne. 2005.

> Imagine a game played at the side of a lake. Two small channels are dug, side by side, leading away from the lake, and the lake water is allowed to fill them up. Part way up each channel, a cork floats, moving up and down with the waves. You stand with your back to the lake and are allowed to look only at the two floating corks. Then you are asked questions about what is happening on the lake. Are there two motorboats on the lake or only one? Is the nearer one going from left to right or right to left? Is the wind blowing? Did something heavy fall into the water? You must answer these questions just by looking at the two corks. This would seem to be an impossible task. Yet consider an exactly analogous problem. As you sit in a room, a lake of air surrounds you. Running off this lake, into your head, are two small channels – your ear canals. At the end of each is a membrane (the ear drum) that acts like the floating corks in the channels running off the lake, moving in and out with the sound waves that hit it. Just as the game at the lakeside offered no information about the happenings on the lake except for the movements of the corks, the sound-producing events in the room can be known by your brain only through the vibrations of your two eardrums. But the sense of hearing finds it easy to answer the same kinds of questions asked at the lakeside: Are there two talkers in the room or only one? Is the nearer one moving from left to right or right to left? Is the kettle hissing? Did something heavy just fall on the floor? These "easy" questions are exactly comparable to the ones asked at the lake. How can the ears answer questions that the eyes cannot? The difficulty is similar in the two cases. The movements of both the corks and the eardrums are determined by the sum of all the waves that enter the channels. The sum is just another wave pattern that does not have, written anywhere on it, that it is actually a sum of a set of waves, or what the component waves might be.
>
> — Albert S Bregman, 2005.[55(p35)]

7.8 Nomenclature

7.8.1 Explanation of Terms

Definitions of terms used in this report.

Term	Definition
acoustically closed earphone:	earphone which is intended to prevent acoustic coupling between the external environment and the ear canal
acoustically open earphone:	earphone which intentionally provides an acoustic path between the external environment and the ear canal
bias:	difference between the expectation of a test result or measurement result and a true value
binaural:	relating to or involving two ears, from the Latin *auris* meaning ear
CI:	Confidence Interval
closed-back earphone:	earphone which does not emit significant sound radiation from the back of the transducer to the external environment
crosstalk:	transfer of signals between communication channels
dBTP	decibels true peak audio level – digital measurement, interpolating between the actual samples in order to take account of over-shoots that would occur later, with, for example, sampling rate conversion
dichotic:	a situation in which the sounds reaching the two ears are not the same
diffuse (sound) field:	sound field in which the effect of the boundary is negligible
diotic:	a situation in which the sounds reaching the two ears are the same
duo-trio test:	see Section 3.1.4 Duo-trio Test

Term	Definition
earphone:	electroacoustic transducer by which acoustic oscillations are obtained from electric signals and intended to be closely coupled acoustically to the ear
experiment:	purposive investigation of a system through selective adjustment of controllable conditions and allocation of resources
free (sound) field:	sound field in which the effect of the boundary is negligible
headphone:	assembly of one or two earphones on a headband or chinband, the use of which may be optional (e.g. with intra-concha earphones)
HRTF:	Head-related transfer function
ILD:	Interaural level difference
ITD:	Interaural time difference
item:	an excerpt, processed by the system under test, a given combination of a programme and codec
localisation:	the perceived location of a sound
LU:	Loudness Units, equivalent to 1 dB, used for differences between loudnesses. e.g. the difference between -23 LUFS and -21 LUFS is “2 LU”
LUFS:	Loudness Units, referenced to digital Full Scale
monaural:	of or involving one ear
MS:	in the context of the crosstalk experiment described in this report, the designated label for the *Mid / Side panning method*: Mid $= L + R$; Side $= L - R$ (also referred to as *Sum and Difference* in the literature)
object:	the system under test, represented by a number of excerpts, which are processed by the system under test

Term	Definition
open-back earphone:	earphone which emits significant sound radiation from the back of the transducer to the external environment
paired comparison:	see Section 3.1.5 Paired Comparison Test
PAN:	in the context of the crosstalk experiment described in this report, the designated label for the *amplitude panning method*: the amplitude panning of a signal between L and R
polarity:	the alignment of a waveform, relative to another waveform, e.g. a device that inverts the polarity of a signal will swap positive values to negative values and vice versa
population:	all those participants that are relevant in a particular study
process:	set of interrelated or interacting activities which transforms inputs to outputs
programme excerpt:	short piece of music, speech or other type of sound event
reference:	test excerpt, reproduced without the processing by a system under test
replication:	repeated application of the same stimulus in order to reduce the standard deviation of the mean value
sample:	any subset of the population of individuals under study
SE:	estimated Standard Error: $s/\sqrt{n}$ (see Section 7.8.2 List of Symbols)
session:	the whole group of trials which are to be evaluated by a subject or a listening panel in a continuous period
stimulus:	reproduction of a certain programme excerpt over a pair of headphones or set of loudspeakers for two channel stereophonic reproduction

Term	Definition
subject:	a test person evaluating the stimuli in a listening test
synthetic signal:	any signal that can be constructed mathematically
triad:	three items given to a subject in the duo-trio test
trial:	subset of a session which begins with the presentation of a set of stimuli and ends with their gradings

7.8.2 List of Symbols

Definitions of symbols used in this report.

α :	Type I error probability (alpha) – also known as false positive rate
β :	Type II error probability (beta) – also known as false negative rate
μ :	Population mean (mu)
Σ :	Summation
σ :	Population standard deviation (sigma)
N :	Number of participants
n :	Number of observations
p:	Probability of success. Also used to indicate a p-value
$\widehat{p}$:	Unbiased estimator of the true probability of success, p
s :	Sample standard deviation, which estimates the population standard deviation σ. Unbiased estimator, assuming random sampling
s^2 :	Sample variance, which estimates the population variance σ^2. Unbiased estimator, assuming random sampling
$\overline{X}$:	Sample mean, which estimates the population mean μ
x :	Observed number of successes
! :	NOT

In accordance with:

European Broadcasting Union. Tech 3343-2016 : Guidelines for Production of Programmes in accordance with R 128. EBU; 2016. https://tech.ebu.ch/docs/tech/tech3343.pdf

International Organization for Standardization. ISO 3534-1:2006 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability. 2006.

International Organization for Standardization. ISO 3534-2:2006 Statistics — Vocabulary and symbols — Part 2: Applied statistics. 2006.

International Organization for Standardization. ISO 3534-3:2013 Statistics — Vocabulary and symbols — Part 3: Design of experiments. 2013.

International Electrotechnical Commission. IEC 60268-7:2011+A1:2020 Sound system equipment — Part 7: Headphones and earphones. 2020.

International Electrotechnical Commission. BS 6840-13 IEC 60268-13:1998 Sound system equipment — Part 13: Listening tests on loudspeakers. 1998.

International Telecommunication Union. ITU-R BS.1116-3 : Methods for the subjective assessment of small impairments in audio systems. ITU-R; 2015. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf

With reference to:

Bech S, Zacharov N. Perceptual Audio Evaluation – Theory, Method and Application. John Wiley & Sons; 2006.

Mason A. WHP 202 Terminology for loudness and level dBTP, LU, and all that. 2011 [accessed 2026 Oct 22]. http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP202.pdf

Wilcox RR. Understanding and applying basic statistical methods using R. John Wiley & Sons; 2017.