SCREENING FOR DIOTIC ACOUSTIC CONTEXT AND HEADPHONES IN ONLINE CROWDSOURCED HEARING STUDIES

Toso Pankovski *[1]
[1]BrainExperiments.com non-profit, Montreal, Quebec, Canada

Résumé

Des preuves expérimentales suggèrent que les expériences en ligne financées par la foule, lorsqu'elles sont adaptées, peuvent produire des données de qualité comparable aux études en laboratoire (Buhrmester, Kwang, Gosling, 2011). L'absence d'une méthode de filtrage fiable pour les casques et le contexte auditif diotique (séparation parfaite des canaux stéréo) est l'une des principales raisons pour lesquelles le crowdsourcing en ligne est rarement possible pour les études auditives. Nous montrons ici que le phénomène des battements parasites peut être utilisé comme méthode de dépistage du contexte diotique avec des résultats satisfaisants. Nous avons recueilli des données par le biais d'une expérience en laboratoire auprès de plus de 2000 participants afin de tester les performances de la méthode par rapport à la référence. Le Kappa de Cohen est de 0,79 (IC 95%, [0,52, 1,06], p<0,001), ce qui donne un "accord substantiel". Les résultats obtenus en laboratoire et en ligne suggèrent que la méthode introduite dans cette étude est adaptée et, par conséquent, qu'elle permet de réaliser des études auditives en ligne basées sur le crowdsourcing.

Mots clefs : dicotique, écouteurs, dépistage, crowdsource, étude en ligne

Abstract

Experimental evidence suggests that crowdsourced online experiments, where suitable, may produce data with quality comparable to in-lab studies (Buhrmester, Kwang, Gosling, 2011). The absence of a reliable screening method for headphones and diotic auditory context (perfect separation of the stereo channels) is one of the main reasons why online crowdsourcing is rarely possible for auditory studies. Here we show that the interference beating phenomenon can be used as a screening method for diotic context with satisfactory results. We collected data through an in-lab experiment from over 2000 participants to test the method's performance against the reference, achieving Cohen's Kappa of 0.79 (95% CI, [0.52, 1.06], p<0.001), yielding "Substantial agreement". The in-lab and online results suggest that the method introduced in this study is suitable, and therefore, an enabler of auditory online crowdsourced studies.

Keywords: diotic, headphones, screening, crowdsource, online study

1 Introduction

The benefits of the interconnected world enable some scientific research to be conducted online with the help of crowdsourced participants, increasing the ability to collect data with large sample sizes. Such studies cost less than if conducted in-lab. Especially in present years (2020, 2021) when people are isolated due to the COVID 19 pandemic, the ability to perform online research is imperative. Researchers are already using Internet-based services to recruit participants from all around the world, such as Amazon's Mechanical Turk or the advertising Google's AdWords service [1-4]. The findings in [3] and [5] even suggest that the samples collected through Amazon's Mechanical Turk are at least of the same quality and as diverse as those collected through traditional means; of course, this is true only for studies where online experiment are appropriate. Considering that, one would think that auditory studies could largely benefit from using online services. Unfortunately, that is not yet the case because online participants are using hardware devices and software (and the rest of the auditory context) that are not under the researcher's control. Therefore, when the experiment design requires strict diotic auditory context (*diotic* means perfect auditory separation between the stereo channels), the absence of control may allow an introduction of bad data in the sample that cannot be identified as such, causing unwanted bias in the results, as suggested in [4], section *2.5. Bias and other possible issues*.

To continuation of our study on tonal consonance [6] required us perform a large number of online listening experiments, differentiating between those conducted in diotic-, from those conducted in non-diotic conditions. Because in online experiments, the researcher has no control over the participant's playing device, we need a listener's diotic context detection method. The requirement for the the participant to use headphones shows insufficient for various reasons, mainly because the playing device (computer hardware, software, headphones) may be defective, or the "spatial sound" features are mixing the stereo channels [7].

The oposite holds true as well – even without wearing headphones, the participant may be able to detect if the

* Toso.Pankovski@BrainExperiments.com

sound is coming from the left or right speaker, yielding a false positive diotic response from the participant [8].

The inherent doubt in participants' responses collected through online experiments creates a necessity for usage of additional prediction method(s) when validating those responses. More details regarding the insufficiency of these questions can be found in the *Supplementary Information* package [9].

2 Method

2.1 The baseline State Interference (SI) method–basic principles

The method we initially attempted to use is briefly explained in this section. We denote this method as the *Stationary Interference* method (SI) because it uses the interference phenomenon (to differentiate from the *non-stationary* interference, such as the interference beating phenomenon). The full description of the SI method is found in the paper where it is introduced initially [1].

The SI method requires the participants to wear headphones; by agreeing to take part in the experiment, the participants *implicitly* declare that they are wearing headphones (as a response to the unasked question: "Are you wearing headphones?"). The implicit response to this implicit question is used as a first predictor of the participant's diotic auditory context - *SI's Declarative* (SI_D) predictor.

The authors of [1] introduced an additional applied acoustics predictor - *SI Acoustic* (SI_A) predictor. In essence, the SI method uses the difference of the perceived loudness between S_0 and S_π sounds presented diotically or not. After obtaining the results from both predictors, the SI method compares them, and in case of a mismatch, it disqualifies the data record. Otherwise, it accepts the result.

2.2 The new BI method–basic principles

The screening method for a diotic acoustical context introduced in this paper is inspired by the baseline SI method. It attempts to mitigate the causes of the SI method's unsatisfactory performance (that we speculated), by improving the accuracy of the SI_D predictor, introducing the new *BI Declarative* (BI_D) predictor, and by improving the accuracy of the SI_A predictor, introducing the new *BI Acoustic* (BI_A) predictor.

Similarly as with the SI method, the BI method compares the results predicted by its two predictors; if they are equal, the method considers the predictions accurate. Otherwise, it disqualifies the data record.

As opposed to the SI method, which classifies the data in two classes: diotic and disqualified, the BI method attempts to classify the data in three classes: diotic, non-diotic, and disqualified. Also, the BI method explicitly asks the participants if they are wearing headphones, as opposed to the SI method, which instructs the participants to wear headphones, but it does not ask them. The responses are collected in initial and interim variables, and later after certain disqualifications, only final list of variables is used (see Table 2).

Abbreviations

BI: Beating Interference – a screening method for diotic auditory context, introduced in this paper

SI: Stationary Interference – a screening method for diotic auditory context (of other authors – a baseline)

BI_D: Beating Interference Declarative Predictor – a predictor used by the BI method, based on the participant's responses to the questions if they are wearing headphones, and the channel separation questions

SI_D: Stationary Interference Declarative Predictor - a predictor used by the SI method, based solely on participants' response to the question if they are wearing headphones

BI_A: Beating Interference Acoustic Predictor – a predictor used by the BI method, based on perception cues produced by beating interference acoustic causes

SI_A: Stationary Interference Acoustic Predictor – a predictor used by the SI method, based on perception cues produced by stationary interference acoustic causes.

2.3 The website of the BI method

The BI method is designed for online listening studies. Therefore it is contained within a website. A demo of the BI method can be found at this website (also, the interference beatings "shadow" stationary zones) [10]. By reviewing the demos, the reader may facilitate her/his understanding of the method.

All the sounds used in the following steps have the following common properties: mp3 file format, 1000 ms duration, 48k Hz sampling frequency, linear ramp-up, and ramp-down, each with a duration of 50 ms. For the in-lab batch, we presented the sounds within a 40 dB SPL quasi-white noise in an attempt to simulate everyday home sounds that the on-line participant would usually face [11]. Their particular properties are specified in the corresponding steps described below.

Step 1: Initial instructions

After accepting the participation through the consent form, the participant is faced with the initial instructions. Important instruction is that the participant is not moving the head during the experiment. One reason for this instruction is to avoid additional bias [12].

Step 2: Channel separation (*ChanLeft* and *ChanRight*)

This step presents two questions, inspecting the volume balance across the left and the right sound channels, and the channel separation. The resultant response is stored in the variables *ChanLeft* and *ChanRight*.

Table 1: The parameters of the used sounds and the names of the related variables that contains the response values

Related response and variable name	Frequency (L:left, R:right)	Mono or Stereo	Presentation level inlab		Additional note
			Diotic	Non-diotic	
A: *OmniEven*	540 Hz	Mono	70 dB SPL	80 dB SPL	
B: *OmniRinging*	500 Hz	Mono	70 dB SPL	80 dB SPL	Trully amplitude modulated 20 Hz
C: *Mut500HzA*	L:500 Hz, R:530 Hz	Stereo	70 dB SPL	80 dB SPL	Diotic: even, non-diotic: ringing
D: *Mut500HzB*					
E: *Mut1k8Hz*	L:1800 Hz, R:1835 Hz	Stereo	60 dB SPL	70 dB SPL	Diotic: even, non-diotic: ringing
F: *Mut3kHz*	L:3000 Hz, R:3035 Hz	Stereo	40 dB SPL	50 dB SPL	Diotic: even, non-diotic: ringing

Table 2: The list of final variables. They are derived from the interim variables, describing data records that are cleaned up from the values which are causing the initial disqualifications. Therefore, the values presented in this table are only those that can be found in the non-disqualified data records. Note that all of the final variables are binary.

Variable Identifier	Variable Description	Value	Value Description
ChanLeft, ChanRight	Declared perceived stereo channel separation for the left-channel-only and right-channel-only sounds, respectively.	1	Perfect Separation
		0	Channels Mixed
Mut500HzA, Mut500HzB	Declared preceived sensation for the same stereo sound QUESTION-500L530R.mp3, presented twice.	1	Even
		0	Ringing
Mut1k8Hz	Declared preceived sensation for the stereo sound QUESTION-1800L1835R.mp3	1	Even
		0	Ringing
Mut3kHz	Declared preceived sensation for the stereo sound QUESTION-3000L3035R.mp3	1	Even
		0	Ringing
WearHp	Declared wearing headphones situation	1	Wearing headphones
		0	Not Wearing headphones
SI, SI_A, SI_D, BI, BI_A, BI_D	Predicted dichotic value by the respective predictors and methods	1	Diotic
		0	Non-diotic

Step 3: Interference beating

- **Participant training phase:** In this step, the website teaches the participant to the meaning of the terms *even* and *ringing* by presenting two sound specially designed for this purpose, one sounding as *even* (*OmniEven* - the interference beatings are not audible) and the other *ringing* (*OmniRinging* - the interference beatings are audible). The perception of *even* and *ringing* are independent if the context is diotic or not (see Table 1).
- **Response collections phase:** The participant is invited to play a sequence of sounds, and to classify each sound as *even* or *ringing*, as per the training obtained in the previous step. The following responses are collected: A, B, C, D, E and F (see Table 1).

Step 4: "Are you wearing headphones?" (variable WearHp)

In this step, the participant is declaring if she/he is wearing headphones while performing the experiment.

Step 5: Speaker or headphones type (variable *SpkType*)

Options with types of speakers and headphones are presented as stylized images. Each image contains a caption text advising if that choice means wearing headphones. In case the participant accidentally made a mistake in the previous step, this could help her/him to realize the mistake. This step offers the participant a choice to change her/his "wearing headphones" response from the previous step.

2.4 Initial disqualification

At this moment, the software has all the data it needs to calculate the predictions. The following quality assurance criteria are used to identify bad data records that should be disqualified:

- If at least one of the channel balance (separation) questions is *silent* (indicating a defective playing device, or significant hearing difficulties), disqualify (variables *ChanLeft* and *ChanRight*).
- If the response to the question "A" (see *Step 3*) is *ringing*, disqualify (variable *OmniEven*).
- If the response to the question "B" is *even*, disqualify (variable *OmniRinging*).
- If the responses to the questions "*Are you wearing headphones?*". (variable *WearHp*) and "*Speaker or headphones type*" (variable *SpkType*) are conflicting (for example, the participant declared that she/he is not wearing headphones, but to the type of headphones question she/he responded with *earbuds*), disqualify (variable *WearHpVSSpkType*).

2.5 The BI Declarative (BI_D) predictor

BI's declarative predictor (BI_D) uses the SI_D predictor in its core, additionally acknowledging that the participant's response to the question "Are you wearing headphones" may be not 100% accurate. Therefore, the BI_D predictor additionally considers the participant's responses regarding the separation of the stereo channels (*ChanLeft* and *ChanRight*). Equation (1) shows the model the BI_D predictor uses to produce its result as diotic or non-diotic:

$$BI_D = \begin{cases} 1, & WearHp + ChanLeft + ChanRight = 3 \\ 0, & WearHp + ChanLeft + ChanRight < 3 \end{cases} \quad (1)$$

where the variables are described in Table 2.

More details can be found in the Supplemental Information package [9]. Here we only assert that this predictor attempts to overcome the previously mentioned false positive diotic responses, where the non-diotic participant are able to determine the exact direction of the sound (left or right) (section 5.1 *Determining the Direction of a Sound Source* in [13]).

2.6 The BI Acoustic (BI_A) predictor

The BI_A predictor uses a stereo sound with perfectly separated left and right channels, each playing pure tones with frequencies F_1, and F_2, satisfying the constraint: $|F_1 - F_2| < 40\ \mathrm{Hz}$. Similarly, as with the SI method, when listening diotically, there is no interference, and the participant should perceive the sound as *even* (the volume of the sound is not changing in time). If the participant is listening to the sound non-diotically, she/he will experience interference beating [13, 14]. In this study, the *beating* sound is denoted as *ringing*. Hence, the particpant's task is to declare if she/he has perceived the sound as *even* or *ringing*.

Equation (2) discloses the model used to calculate the predicted diotic context based on the BI_A; if we define *SumOfMut* as:

$$SumOfMut = Mut500\mathrm{Hz}A + Mut500\mathrm{Hz}B + Mut1k8\mathrm{Hz} + Mut3k\mathrm{Hz}\,;$$

then for the BI_A predictor result we have:

$$BI\_A = \begin{cases} 1, & SumOfMut \geq 3 \\ 0, & SumOfMut < 3 \end{cases}; \quad (2)$$

where the variables are described in Table 2.

Similarly to the SI_A predictor [1], the BI_A predictor is using the voting paradigm to derive the resultant prediction. In BI_A's case, a diotic context is predicted if *more than half* (3 or more out of 4) responses to the mutable-amplitude sounds are declared as *even*.

The in-lab preliminary simulations demonstrated the existence of stationary spatial zones where the interference beatings are inaudible ("shadow" zones) – they may affect the accuracy of the BI_A method, yielding a false positive diotic response from the participant. A software simulation in MATLAB has been developed to analyze if these "shadow" zones are produced by the room acoustics (the reflections of the sounds wave from the surrounding objects) – the simulation confirmed the said speculation (see second part of the video [10]). Another phenomenon may additionally impede the BI_A predictor: that is the binaural beating – a phenomenon that causes the participant to experience beatings-like neural sensations, introducing slight bias in the method – false negative diotic responses[13, 15, 16].

2.7 Final disqualification step and result

The predicted value calculated according to the BI_A predictor calculated per Equations (1) and (2) are compared. If these two values are not equal, the BI method is disqualifying the data record, otherwise the result is declared as final.

2.8 Participant, recruitment, and technology

This study is ethics-approved by the Independent Ethics Research Board VeritasIRB (www.VeritasIRB.com, Montreal, Canada). The participation was completely anonymous, and in the online experiments, no information about the participant's age, race, gender, sex, or location were collected (the age and sex information are collected in the in-lab experiment). The participants were not asked if they have hearing problems, or if they are aware of any technical problems with their playing devices they used in the experiment.

In this study, we collected three batches of data samples. The details of the batches are outlines in Table 3. The raw results of the experiments obtained through these batches are located in the Excel files in the Supplementary Information package [9].

Table 3: Descriptions of the participant batches

Batch	Sample Size	Type/Incentive [USD]	Wearing headphones
1	1656	Online/0.15	Choose
2	519	Online/0.15	Asked to wear
3	18	Inlab/0.00	10 weared 8 not weared

The in-lab batch participants provided their demographics (N = 18, 55.6% female, 44.4% male). At the time of the experiment, their mean age was 33.6 years (SD = 13.2), and they resided in Montreal, Canada. All the in-lab participants declared that they have a healthy hearing.

The sounds were generated and edited using the software GoldWave version 6.24. The in-lab testing utilized JVC Over-Ear Headphone HA-RX330 with frequency response 12 Hz-22000 Hz for the experiments performed with headphones, and embedded laptop speakers for the in-lab non-headphones experiments. The delivery of the specific dB SPL levels is achieved only in-lab, by using Proster' digital sound level meter HT80A (40 dBA - 130 dBA; Accuracy +/-1.5 dB; Resolution: 0.1 dBA; frequency range 31.5 Hz-8 kHz; standard applied: IEC 651 type 2, ANSI S1.4 type 2). In-lab, the dB SPL measurement was taken at the ear-position when the participant is sitting in front of the laptop and the sound is delivered by the embedded laptop speakers. The headphones dB SPL in-lab delivery was first subjectively calibrated comparing to the one measured for the laptop speakers and ensured the same volume level is selected for every participant. There was no enforcement of the dB SPL delivery for the online experiments. The "spatial sound" and the automatic subjective loudness equalization features were disabled on the laptop used in-lab.

2.9 Procedure

All the experimental batches (online and in-lab) are collected through the same website and by using the same procedure. After completing the BI method's steps, the participants are asked to perform the baseline SI method, which allows for a cross-method pair-wise comparison of the results. The data was cleaned up by the initial disqualification criteria. Over the initially non-disqualified records the BI_A, BI_D, SI_A, SI_D, SI, BI, and in-lab control (CR) predictions are calculated. As a main statistic the Cohen's Kappa inter-rater agreement between two methods/predictors corrected by chance (see Table 4), and ROC analyses are used [17-21].

3 Results and discussion

A large volume of the raw data, the interim, and the final detail results are located in the Supplementary Information package since they cannot fit in this paper. Here we present only the results we consider most representative to the new BI method and its comparison with the baseline SI method.

Table 4: Cohen's Kappa agreement gradations as per [22]:

Cohen's Kappa Value	Descriptive Gradation
<0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

The results of the in-lab experimental batch (batch 3) are presented in Table 5, where we observe superior results of the BI predictors and BI method over the SI predictors and method, respectively. As a consequence, from the in-lab superiority of the BI over the SI method in-lab (Table 5), in the presentation of the online data (Table 6) the BI method is used as a reference method. Although the SI method does disqualify the non-diotic data records, to fairly compare the methods, in Table 6 we considered a SI method version that does not disqualify the non-diotic data records. The first row of the table shows the most important result, where the SI method classifies 104+43=147 data records as diotic, where the BI method disqualifies them - that means that the SI method declares 8.89% (95% CI, [7.55%, 10.4%]) false diotic results (if we consider the BI's classification absolutely accurate). In addition, the SI method disqualifies 202 data records that are classified by BI as diotic - that is 12.2% (95% CI, [10.7%, 13.9%]) wasted data of the whole sample.

Table 5: Comparison of the performances between the analyzed predictors and screening methods. In order to avoid calculation errors for Cohen's Kappa for lines 5 and 6, additional treatment of the data has been performed (see Supplementary Information package for more details).

				PREDICTION TEST STATISTICS									
				Cohen's Kappa					**Accuracy**				
Line	**Batch**	**Method or Predictor**	**In reference to In-Lab Control (true) value**	**Sample Statistics**	**Arbitrary Interpretation of the obtained agreement**	**95% CI Lower Bound**	**95% CI Upper Bound**	**p-Value**	**Sample Statistics**	**95% CI Lower Bound**	**95% CI Upper Bound**	**p-Value [Acc>NIR]**	**No Information Rate**
1	In-lab	**BI_D**		**0.85**	**Almost perfect**	0.57	1.13	0.002	**0.94**	0.73	1.00	0.023	0.72
2		**SI_D**		**0.47**	**Moderate**	0.08	0.86	0.018	**<u>0.72</u>**	0.47	0.90	*0.618	0.72
3		**BI_A**		**0.87**	**Almost perfect**	0.62	1.12	<0.001	**0.94**	0.73	1.00	0.023	0.72
4		**SI_A**		**0.42**	**Moderate**	-0.02	0.85	0.047	**<u>0.72</u>**	0.47	0.90	*0.618	0.72
5		**BI**		***0.79**	**Substantial**	0.52	1.06	<0.001	**0.89**	0.67	0.99	0.034	0.68
6		**SI**		***0.19**	**Slight**	-0.04	0.42	0.014	**<u>0.30</u>**	0.12	0.54	*1.000	0.70
7		**BI (diotic BI only)**		**1.00**	**Almost perfect**	1.00	1.00	0.001	**1.00**	0.81	1.00	0.011	0.78
8		**SI (diotic SI only)**		**0.85**	**Almost perfect**	0.57	1.13	0.002	**0.94**	0.73	1.00	*0.067	0.78

Table 6: Confusion matrix of the Stationary Interference (SI) baseline method, where the new Beating Interference (BI) method is used as a reference standard. This table shows where the SI method classifications differ (or fails, considering Table 5 results) from those of the BI method. To produce this table, we considered an SI method version that does not disqualify the non-diotic data records. The results are obtained from the online Batch 1.

		Beating Interference (BI) Method			
		Diotic	**Non-diotic**	**Initially disqualified**	**Finally disqualified**
Stationary Interference (SI) Method	**Diotic**	**421**	**0**	104	43
	Non-diotic	**61**	**328**	58	113
	Initially disqualified	0	0	0	0
	Finally disqualified	202	100	125	101

Finally, the results obtained from the online Batch 2, where the participants are asked to obligatory wear headphones (same as the original SI study), are used to compare the accuracy of our implementation of the SI method, to its original implementation and the design purpose as described in [1]. As such a measure, we use single statistic: the total disqualification proportion obtained as per the SI method. We obtained an SI disqualification proportion of 37.0% (95% CI, [32.83%, 41.31%]), which is close to the disqualification proportion of 35.3% reported by [1]. The proximity of these two statistics provides confidence that this study replicated the SI method sufficiently accurately.

Lastly, we decided not to use the psychophysics methods of binaural masking level difference (BMLD) and interaural time-difference discrimination (ILD) due to their small magnitude of differences between the diotically and non-diotically perceived sounds. We have not conducted experiments to confirm these concerns (it is out of the scope of this study); another study may be conducted to investigate the viability of the BMLD and ILD paradigms in uncontrolled online experiments.

Regarding the BI method, its magnitude of amplitude fluctuation (the ringing) is easily noticeable (even in high noise level conditions), as long as the top-amplitude portion of the sound is audible – we consider this as the most beneficial feature of the BI method, which makes it suitable for uncontrolled online auditory studies.

Finally, there are many details regarding the BI method including speculated reasons for BI method superiority are not disclosed in this paper due to length constraints. They can be found in the SI package [9].

4 Conclusion

The results of this study demonstrate that the *Beating Interference* (BI) method can be used as a screening method to determine if the participant's auditory context in online crowdsourced hearing studies is diotic. It uses two distinct predictors to predict the diotic context of the data record, disqualifying those that yield mismatching prediction results. The results of this study are obtained in in-lab and online batches, and the results suggest that the *Beating Interference* (BI) method shows somewhat better performance than the baseline *Stationary Interference* (SI) method [1].

The reader can watch a demo of the *Beating Interference* (BI) diotic screening method by following the enclosed link (also, a video demonstrating the beatings "shadow" zones is enclosed): [10], whereas Supplementary Information is also available [9].

Remerciements/Acknowledgments

Ana Pankovska (certified as *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS2: CORE*) helped by verifying some of the calculations, and with editing the text.
Eva Pankovska helped by proofreading the text.

References

[1] K. J. Woods, M. H. Siegel, J. Traer and J. McDermott, "Headphone screening to facilitate web-based auditory experiments," Atten Percept Psychophys, vol. 79, no. 2064, 2017.

[2] K. A. Thomas and S. Clifford, "Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments," Computers in Human Behavior, vol. 77, pp. 184-197, 2017.

[3] M. Buhrmester, T. Kwang and S. D. Gosling, "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?," Perspectives on Psychological Science, vol. 6, pp. 3-5, 2011.

[4] X. Ma, M. Cackett, L. Park, E. Chien and M. Naaman, "Web-Based VR Experiments Powered by the Crowd," in Proceedings of the 2018 World Wide Web Conference (WWW '18), Republic and Canton of Geneva, Switzerland, 2018.

[5] D. R. Saunders, P. J. Bex and R. L. Woods, "Crowdsourcing a normative natural language dataset: A comparison of Amazon Mechanical Turk and in-lab data collection," Journal of Medical Internet Research, vol. 15, p. e100, 2013.

[6] T. Pankovski and E. Pankovska, "Emergence of the consonance pattern within synaptic weights of a neural network featuring Hebbian neuroplasticity," Biol. Insp. Cong. Arch., vol. 22, pp. 82-94, 2017.

[7] T. Lee, Y. Baek, Y. C. Park and D. H. Young, "Stereo upmix-based binaural auralization for mobile devices," IEEE Trans. Consum. Electron, vol. 60, no. 3, pp. 411-419, 2014.

[8] J. Blauert, Spatial Hearing - The Psychophysics of Human Sound Localization, Cambridge, MA; London, England: The MIT Press, 1983, p. 148.

[9] T. Pankovski, "Supplementary Information - Screening for Dichotic Acoustic Context and Headphones in Online

Crowdsourced Hearing Studies," 22 April 2020. [Online]. Available: https://figshare.com/s/e9818607ded7dcad2ff4. [Accessed 03 March 2021].

[10] T. Pankovski, "Video Demo - BrainExperiments.com," 2020. [Online]. Available: http://brainexperiments.com/headphonesVIDEO.aspx. [Accessed 03 March 2021].

[11] "Loud Noise Can Cause Hearing Loss," Centers for Disease Control and Prevention, [Online]. Available: https://www.cdc.gov/nceh/hearing_loss/what_noises_cause_hearing_loss.html. [Accessed 25 June 2019].

[12] B. M. Salin, O. N. Kemarskaya and M. B. Salin, "Estimating the Amplitude of Wind Waves by Correlation Analysis of the Reverberation Signal," Acoustical Physics, vol. 65, no. 1, p. 41–51, 2019.

[13] J. Scnupp, I. Nelken and A. King, Auditory neuroscience: making sense of sound, Cambridge, Massachusetts: MIT Press, 2011.

[14] "Interference beats and Tartini tones, Physclips," 2 July 2018. [Online]. Available: http://www.animations.physics.unsw.edu.au/jw/beats.htm.

[15] J. J. Groen, "Super- and Subliminal Binaural Beats," Acta Oto-Laryngologica, vol. 57, no. 3-6, pp. 224-230, 1964.

[16] D. McFadden and E. G. Pasanen, "Binaural beats at high frequencies," Science, vol. 190, no. 4212, pp. 394-396, 1975.

[17] J. Cohen, "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37-46, 1960.

[18] W. Tang, J. Hu, H. Zhang, P. Wu and H. He, "Kappa coefficient: a popular measure of rater agreement," vol. 27, no. 1, pp. 62-67, 2015.

[19] A. DiCenso, G. Guyatt and D. Ciliska, "A Guide to Clinical Practice," in Evidence-Based Nursing, Elsevier MOSBY, 2005.

[20] P. F. Watson and A. Petrie, "Method agreement analysis: A review of correct methodology," Theriogenology, vol. 73, p. 1167–1179, 2010.

[21] M. J. Warrens, "A COMPARISON OF COHEN'S KAPPA AND AGREEMENT COEFFICIENTS BY CORRADO GINI," International Journal of Recent Research and Applied Studies, vol. 16, no. 3, 2013.

[22] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," Biometrics, vol. 33, pp. 159-174, 1977.