Design of an Interactive Earphone Simulator and Results from a Perceptual Experiment

PerMagnus Lindborg
Nanyang Technology University
permagnus@ntu.edu.sg

Miracle Lim Jia Yi
Nanyang Technological University
miraclejoseph.lim@gmail.com

ABSTRACT

The article outlines a psychoacoustically founded method to describe the acoustic performance of earphones in two dimensions, *Spectral Shape* and *Stereo Image Coherence.* In a test set of 14 typical earphones, these dimensions explained 66.2% of total variability in 11 acoustic features based on Bark band energy distribution. We designed an interactive *Earphone Simulator* software that allows smooth interpolation between measured earphones, and employed it in a controlled experiment (N=30). Results showed that the preferred ‘virtual earphone’ sound was different between two test conditions, silence and commuter noise, both in terms of gain level and spectral shape. We discuss possible development of the simulator design for use in perceptual research as well as in commercial applications.

1. INTRODUCTION

One of the most common situations for music consumption today might very well be that of listening over earphones while on a suburban train or bus during rush hours. The acoustic performance of commercially available earphones is highly variable, and it is not clear to what extent objective audio quality measures predict people’s preference in a given listening context. Portable audiovisual entertainment devices are increasingly popular and sales figures of earphones have increased exponentially within the past decade. According to *Cellular-news*, combined headphone and earphone sales in Southeast Asia went up by 7% during the first half of 2010 alone, and even more so in Singapore [1]. The growing demand is one indicator of the direction in which technology is changing lifestyle and habits in the early 21st century.

This is the background for a project to investigate the perceived quality of earphones. A questionnaire survey of listening habits of commuters on public transport in Singapore was conducted (N=94). Among other things, it revealed that people use earphones in a wide price range: from ‘free’ (e.g. included with player or phone) to several hundred dollars worth. Results showed a positive relationship between cost and perceived quality. However, we suspected a less direct relationship with objective audio quality, itself a multidimensional measure that would have to be calculated from acoustic features.

Fourteen earphones were selected, with characteristics typical of those observed in the survey findings, and their acoustic performance was measured in studio. A controlled experiment with volunteers was designed to determine perceptual ratings of sound quality, as well as visual aesthetics, physical comfort, and perceived sound quality in conditions of ‘lab silence’ and ambient noise. To achieve a high degree of ecologic validity, we used in the noise condition actual soundscape recordings from a commuter train, reproduced at the SPL that was registered on-site.

1. AIMS FOR THE SIMULATION

To be able to make predictions of earphone sound quality ratings, we developed an interactive earphone simulator to be part of the experiment. The design was made in order to minimise bias and to let the person doing the ratings quickly find the preferred ‘virtual earphone’ sound in a given condition, i.e. in a noisy environment or in lab silence.

It has been shown that perceptual ratings of subjective features are correlated with loudness level. In a real-life situation, such as listening to music while commuting, the user adjusts for optimal loudness considering factors such as the kind of sound (e.g. music style), the internal emotional state and cognitive attitude, while taking into account the level of noise in the prevailing sonic environment. As a consequence, in an experimental setting, the user must be allowed to adjust the playback gain for optimal experience when shifting between different earphones. The trivial observation about actual usage also implies that SPL on its own is not a meaningful feature for earphone acoustic performance. Therefore, we hypothesised that frequency magnitude response and stereo image would be sufficient to describe earphone sound quality.

For the research project as a whole, several other acoustic features were considered, i.e. noise isolation, harmonic distorsion, and impedance matching, as well as non-acoustic features such as physical comfort, visual aesthetic, and price. The results are reported in [4]. How multimodal perceptual features relate to objective acoustic features is discussed in [9] and goes beyond the scope of the present text.

In what follows, we first describe how the acoustic measurements were made. Then, the design of an interactive *Earphone Simulator* and an implementation using the acoustic measurements. Finally, we report results from a controlled pilot experiment (N=30).

2. ACOUSTIC MEASUREMENTS

Table 1 lists the selection of 14 commercially available earphones, representative of those typically used by commuters on buses and trains in Singapore. Purchase prices were in a range from zero ('free') to around 400 USD. Four use buds placed in the outer ear, and ten use in-ear buds of different shape and material, such as foam, smooth silicon, and 'tree' shaped silicon.

1. Procedure

Measurements were made in accordance with best practices as in [6]. Impulse-response recordings were conducted in an acoustically isolated sound booth. A time-smoothed impulse or 'swept sine wave' (logarithmic, 30 seconds) was generated. The frequency range 12…22050 Hz was chosen in order to cover the defined range of Bark bands. Recall that the lower limit of band 1 is 50 Hz, and the higher limit of band 24 is 15500 Hz. The chirp was played back via a sound card (*Echo AudioFire4*) through one earphone at a time, with earbuds fitted in left and right pinnae of a manikin head (*Neumann KU100*). Left and right responses were captured by built-in reference microphones. A total of 33 stereo recordings were made of the 14 earphones, with left-right swapping of earphone buds in the manikin pinnae to minimise any bias introduced by frequency response mismatch between the microphones. Custom software developed in *Max* (*Cycling'74*) was then used to calculate each channel's energy content in 24 Bark bands [2], [3]. Plots of the earphone responses in units of Bark band are shown in Figure 1.

2. Results

Numerous features of the profiles were investigated before a parsimonious set of features could be settled upon. Seven measures of frequency magnitude response were calculated on the response averaged across left and right channels. Note that the relation between levels in broad Bark band regions and the total SPL is a measure of spectral shape. Means were calculated on amplitude, i.e. linear pressure equivalent, while slope was calculated on levels expressed on a decibel scale [5].

- *BB_pki* = index for the Bark band with highest level;
- *SPL_low* = mean of Bark bands 1…8 minus total SPL;
- *SPL_mid* = mean of bands 9…16 minus total SPL;
- *SPL_high* = mean of bands 17…24 minus total SPL;
- *R_low* = regression slope (Pearson's *r*) of bands 1…8;
- *R_mid* = slope of bands 9…16; and
- *R_high* = slope of bands 17…24.

Four measures of left/right channel matching were calculated on the separate response of left and right channel. Note that the correlation *r* between responses was considered but not included in the final selection.

- *ChD_rms* = root mean square of channel differences;
- *ChD_low* = RMS of differences in bands 1…8;
- *ChD_mid* = RMS of differences in bands 9…16; and
- *ChD_high* = RMS of differences in bands 17…24.

Numeric values for these measures are listed in Table 1.

3. Analysis

The interrelationships of the features were investigated with a Principal Component Analysis approach. The first two components together explain 66.2% of the variability in the data. The original solution was rotated so as produce two derived dimensions whose meaning could easily be interpreted. The first axis, explaining 43.0%, describes *Spectral Shape*: low values correspond to earphones with 'boomy' sound, and high values to those with 'brighter' character. The second axis, explaining 23.2%, describes *Stereo Image Coherence*: low values mean that left and right channels have differing Bark band profiles, and high values that responses are closely matching. Each earphone thus occupies a position in a plane with orthogonal axes. Figure 2 shows a biplot of the rotated PCA.

3. AN EARPHONE SIMULATOR

A software simulation was designed to enable participants in the ensuing perceptual experiment to interactively select their preferred 'virtual earphone' sound.

1. Interpolation space

Each of the 14 measured earphones is represented by an {x, y} position, or node, in the plane with axes corresponding to *Spatial Shape* and *Stereo Image Coherence*, i.e. the two rotated PC dimensions. The Bark band left/right profiles of an intermediate point in this plane can be estimated as a linear interpolation of values from two or more fixed positions weighted by the inverse of their Euclidian distance to that point. The design was implemented in a Max patcher, using *FTM* [7] to store 51 values for each earphone, i.e. name, PC-derived position, and measured frequency response levels in 24 Bark bands per channel. The size of the region within which an earphone measurement contributes to an interpolation must be decided. Because the 14 measurements are not equally distributed in the plane we have defined, the size of the region around some nodes must be extended so as to achieve smooth interpolations and minimise non-covered space. A solution was found heuristically where each region is a circle with radius adjusted so as to cover the two closest neighbours and exactly touch the third. The interpolation space is visualised in Figure 3.

Table 1. Values of the selected acoustic measures of 14 earphones.

	BB_pki	SPL_low	SPL_mid	SPL_high	R_low	R_mid	R_high	ChD_rms	ChD_low	ChD_mid	ChD_high	PC1	PC2
A	15	-24.56	-5.88	-7.27	0.86	0.94	-0.84	1.66	0.99	2.50	1.03	1.86	1.54
B	15	-16.13	-5.79	-9.62	0.95	0.72	-0.92	5.37	6.83	3.54	5.22	0.89	-0.51
C	4	-8.41	-10.31	-10.04	0.64	-0.95	-0.93	2.58	2.84	1.65	3.04	-1.66	0.54
D	16	-9.51	-7.66	-11.99	0.41	0.37	-0.94	3.35	2.60	2.17	4.71	-0.59	0.61
E	12	-20.06	-4.04	-11.29	1.00	-0.07	-0.70	1.44	2.11	0.86	1.03	1.30	1.13
F	19	-9.42	-9.49	-9.72	0.71	-0.93	-0.64	3.70	4.86	3.66	2.01	0.33	-1.09
G	16	-19.36	-6.59	-7.45	0.99	0.95	-0.93	3.75	5.62	2.18	2.42	1.50	0.59
H	22	-22.71	-9.95	-4.31	0.97	0.99	-0.60	4.15	5.16	4.91	1.03	2.67	-0.33
I	1	-4.52	-12.41	-15.61	-0.99	0.72	-0.97	1.27	0.84	0.69	1.92	-4.03	1.85
J1	14	-8.90	-7.44	-13.29	0.64	0.82	-0.92	1.33	0.91	0.63	2.02	-0.51	1.69
J2	14	-9.86	-6.78	-13.14	0.98	0.47	-0.92	7.78	12.71	1.43	4.24	0.10	-2.48
J3	14	-23.32	-4.86	-8.86	0.96	0.88	-0.91	1.74	1.64	0.83	2.40	1.56	1.66
J4	17	-10.00	-6.95	-12.59	0.81	0.83	-0.97	4.03	6.55	0.67	2.31	0.14	0.10
K	3	-5.87	-9.53	-16.07	-0.72	-0.12	-0.73	13.23	10.20	13.31	15.61	-3.55	-5.28

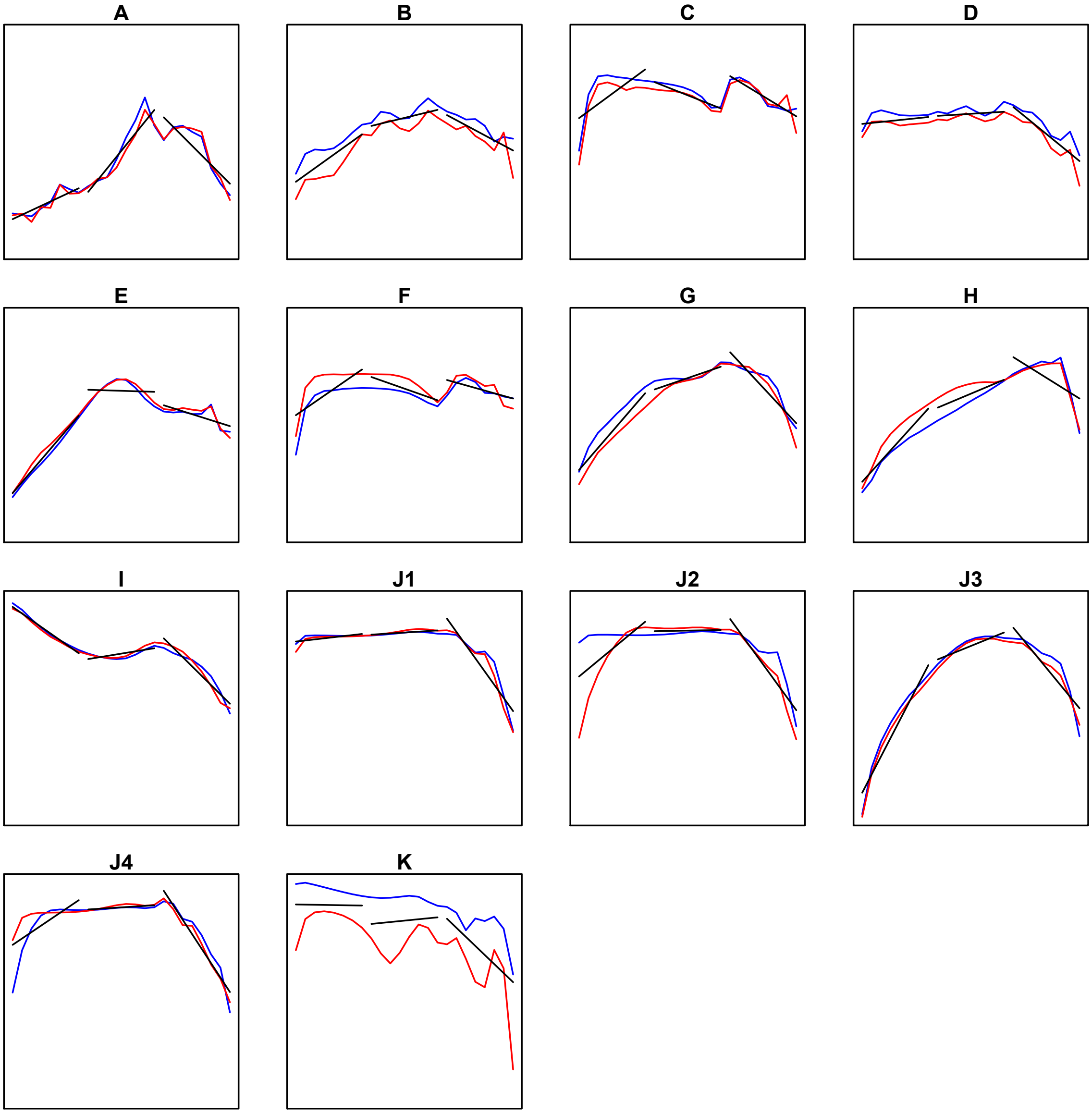

Figure 1. Averaged frequency responses of 14 earphones in 24 Bark bands for left (blue) and right (red) channels. Linear regression lines (‘slopes’, black) are indicated for channel average (‘mono mix’) in low, mid, and treble Bark band ranges.

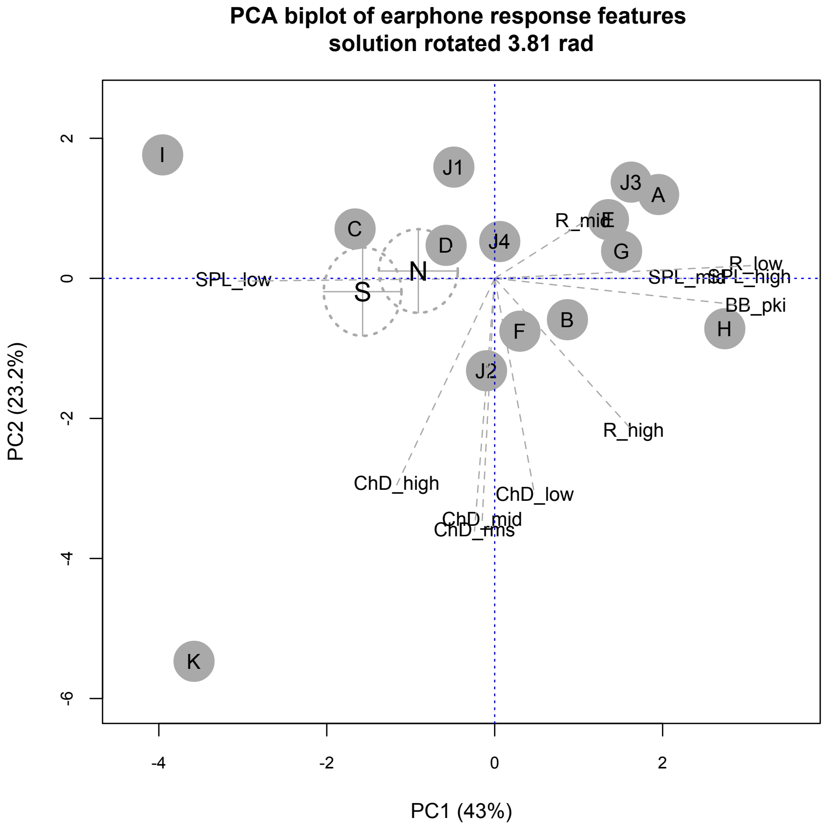

Figure 2. PCA biplot of 14 earphones where 11 acoustic measurements are projected onto a plane with axes *Spectral Shape* and *Stereo Image Coherence*. 'S' and 'N' refer to the preferred virtual earphone sound in *Silent* and *Noise* conditions (mean position across 30 participants, surrounded by 95% confidence ellipses).

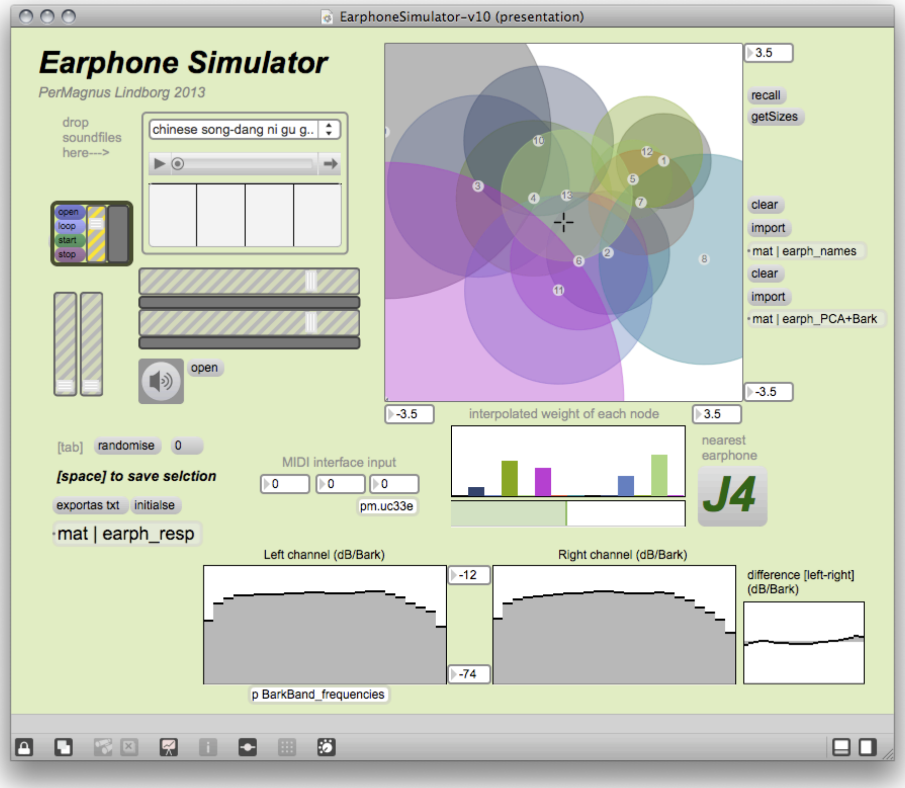

Figure 3. User interface for the Earphone Simulation. The square with the colourful circles corresponds exactly to the 2-dimensional plane yielded by the PCA. It is a visualisation of the space used for interpolation of Bark band profiles.

2. Slider interface

The patcher receives interactive input from three sliders of a USB MIDI interface (*Evolution UC33e*). One slider allows the adjustment of gain level, for reasons discussed above. The two other sliders are mapped to the PCA-derived dimensions, in a particular way. When using a physical input device to represent some perceptual dimension, there might be tacit assumptions such as "increased pitch goes upwards" or "increasing loudness goes rightwards", and so forth. Such 'cliché' mappings can introduce response bias. To reduce this bias, the design randomises slider mappings (input slider -> x or y axis) as well as whether sliders and axes are mapped straight, or mirrored. This gives eight different arrangements. Since the mapping is not communicated to the person doing the rating, she must inform herself through attentive listening while moving the sliders of the interface. For each rating stimulus, e.g. a sound excerpt or a new condition, the mapping is randomised. Further, in order to insure against 'lazy clicking' bias, the software verifies that a certain amount, i.e. at least 10 out of 14 nodes have been 'heard' (have been part of an interpolation), before a preference rating is accepted and saved to disk.

3. Filterbank

The interpolated 2x24 values, determining the Bark band profiles of the channels of a 'virtual earphone', are sent to a filter bank, implemented as a set of parallel 3rd-order Butterworth bandpass filters with centre frequency and bandwidth as in [2], [3]. The user can thus smoothly move between different kinds of earphone filtering, and eventually select what s/he consideres the optimal sound. In the experiment, reference earphones with a very flat frequency response were used (*Etymotic ER-2*). Compared to the commercial earphones that are simulated, these earphones can be considered transparent. According to the manufacturer, their passive isolation with 'shallow insertion'. Finally, the Bark profiles for the preferred sound are saved to disk, together with the amount of gain adjustment for playback level. The *Earphone Simulator* interface is shown in Figure 3.

4. EXPERIMENT

The software was employed in a controlled experiment, as part of a pilot research project to investigate several aspects of earphones [4].

1. Procedure

30 volunteers completed the experiment, one at a time, taking approximately 10 minutes of the test session. Participants received a movie voucher as a token of appreciation. The participant was fitted with the set of reference earphones (*ER-2)* and presented with the interface (*UC33e)*. As stimuli, songs were selected randomly from a collection that had been normalised in terms of RMS. The participant was informed that two sliders control "the sound" (but not in what way) and that one slider controls "the volume". There were two conditions, presented in random order. The 'Lab Silence' condition (ie. the sound studio) was measured at Leq(A, 60s)=39.8 dB, Leq(C)=65.4 dB. The 'Commuter Noise' condition, where a recording from the interior of a Singaporean MRT train during rush hour was played back at the level registered at the original site, was in studio measured at Leq(A)=75.7 dB, Leq(C)=82.6 dB. Hence the difference in ambient noise level between conditions was substantial. The participant's task was to move the three interface sliders so as to select the "best sound" for the given condition. They repeated the task 6 times or more for each condition, and were free to change songs at any time. As described above, the mapping of slider movement to PCA dimension changed randomly between 8 different configurations every time a new song was selected. This obliged the participant to listen out carefully for how the sliders affected the sound output. To sum up, three parameters determining the preferred virtual earphone were collected. They are here referred to as *Spectral Shape*, *Stereo Image Coherence*, and *Level*. The first two are identical to the {x, y} position in the 2-dimensional rotated PCA plane, described above.

In the first round (N_1=13) a procedure problem caused gain levels to be incorrectly saved. Serendipitously, screenshots had been taken of the GUI for all participants preferred setting in either condition, and in several cases for both conditions. From the latter, correct gain adjustments could be read directly, and for the remaining, reasonable estimates could be inferred with a conservative *ad hoc* method. As a result, *Level* values were similar to those in round two (very carefully registered), but because of the conservative estimate made, they showed a less pronounced difference between conditions.

2. Results

Means (on linear pressure equivalent where appropriate) were calculated for each participant and condition. A repeated-measures MANOVA with *Spectral Shape*, *Stereo Image Coherence*, and *Level* as dependent variables, and *Condition* as independent variable, yielded the results in Table 2.

Table 2. Main results from repeated-measures MANOVA of Condition onto 3 parameters of the preferred virtual earphone. Cohen's *d* uses the pooled standard deviation method.

variable	F(1, 29)	p	*d*	ω^2
Spectral Shape	7.32	0.0113 *	0.531	0.196
St. Img. Coherence	0.580	0.452	0.179	0.019
Level	11.7	0.0019 **	0.667	0.280

As expected, *Level* was clearly different between conditions. It was on average 4.2 dB higher during the noise condition, with 95% confidence interval {3.2…7.8} dB. The effect size was two-thirds of a standard deviation, and *Condition* explained 28% of the variance in *Level.* Interestingly, there was a significant difference in *Spectral Shape* between conditions. During the noise condition, participants preferred an earphone sound with larger ratio between higher Bark bands energy to lower Bark bands energy, i.e. *SPL_high* divided by *SPL_low;* see Section 2.2, and Figure 4. Given that the commuter train sonic environments (e.g. the recording used in the noise condition) contains a lot of low-frequency energy, music would be heard more clearly through an earphone with a high-frequency spectral bias. The effect size was slightly more than half a standard deviation, and *Condition* explained nearly 20% of the variance in *Spectral Shape.* For *Stereo Image Coherence* the difference between conditions was not significant.

In Figure 2 the positions of the optimal (preferred) virtual earphone sound can be seen, in both conditions ('S'=*Lab Silence*, 'N'=*Commuter Noise)*. Note that neither corresponds exactly with the sound profile of any of the 14 measured earphones.

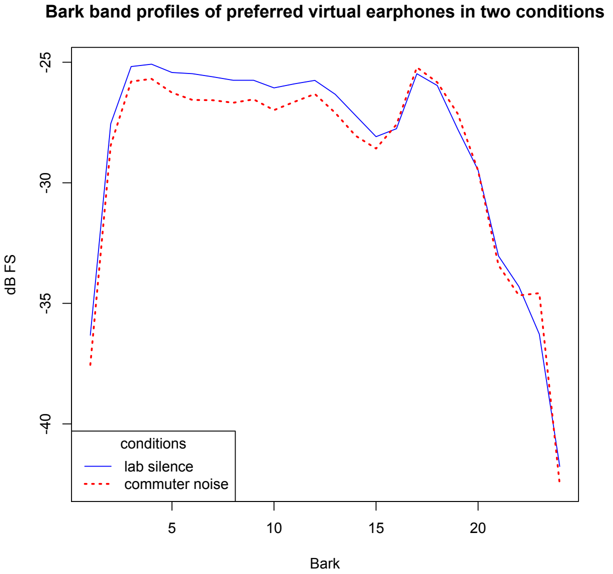

Figure 4. Bark band profiles of preferred virtual earphones (mean across participants). Blue line is in the 'lab silence' condition, and dashed red line is in the 'commuter noise' condition.

3. CONCLUSION

We have described a psychoacoustically founded method to analyse acoustic measurements of earphones and the design of a prototype *Earphone Simulator* software. A pilot experiment employing the Simulator showed that, in addition to (gain) *Level, Spectral Shape* was a useful dimension along which listeners differentiated their preferred sound under two ambient noise conditions.

One reviewer of this article brought to our attention recent work [8] by Rämö and Välimäki on software simulation of headphones in quiet and noisy situations. The authors measured the frequency response and isolation capabilities of different headphones. This is a highly interesting work that merits further study, in particular in regards to the reference headphone calibration method and the inclusion of noise isolation in the simulation software. We believe that the Earphone Simulator described in this article has features that are not described in their work, in particular the possibility to create a 'virtual earphone sound' by smooth linear interpolation between measured, real-life earphones and using a physical interface. One interesting avenue of future work could be to integrate the methods in [8] with those we have presented here.

We believe that interactive simulations enable certain kinds of perceptual investigation and that they extendable. Further development could aim to integrate all parts of the method here described in a single software, i.e. impulse-response measurements, PCA, interactive interpolation, and perceptual ratings. Such a software would be adaptable to various research design scenarios involving perceptual ratings of earphones, headphones, or loudspeakers of any type. It would also potentially be valuable in a commercial situation where a user needs to make an optimal selection within a set of loudspeaker options, depending on personal preferences of sound quality as well as other factors.

4. ACKNOWLEDGMENTS

The 1st author is the supervisor of the 2nd author's undergraduate Final Year Project. The work here reported mainly concerns the contribution of the 1st author and represents only a part of the research work undertaken for the FYP. We thank the anonymous SMAC-SMC reviewers for their valuable comments.

5. REFERENCES

[1] Cellular-news. (2011) "Increased Demand for More Sophisticated Headphones, Earphones, and Headsets". http://www.cellular-news.com/story/51102.php (April 2013)

[2] Zwicker, E. and Fastl, H. (1990, -9). *Psychoacoustics Facts and Models*. Springer-Verlag, Munich Germany (2nd edition).

[3] Loy, G. (2007). "Psychophysical Basis of Sound". Chapter 6 in *Musimathics. The Mathematical Foundations of Music.* Vol. 1 & 2. MIT Press, England.

[4] Lim, M. J. Y. & Lindborg, PM. (2013). "How much does Quality Matter? Listening over earphones on Buses and Trains". *Proc. ICME3*. Jyväskylä, Finland.

[5] British Standards Institution (2007). *Acoustics - Definitions of Basic Quantities and Terms.* BSi ISO/TR 25417:2007.

[6] British Standards Institution (2009). *Acoustics - Measurements of room acoustic parameters.* BSi ISO 3382-1:2009.

[7] FTM & Co. http://ftm.ircam.fr/ (April 2013).

[8] Rämö, J. & Välimäki, V. (2012). "Signal Processing Framework for Virtual Headphone Listening Tests in a Noisy Environment". *132nd AES Convention*.

[9] Lindborg, PerMagnus (2013, submitted). "Perception of soundscapes correlates with acoustic features and is moderated by personality traits."