Perception of sound quality in mobile devices is affected by device type and usage context

Maria Uther
University of Winchester
Winchester, United Kingdom, SO22 4NR
Maria.Uther@winchester.ac.uk

Adrian Banks
University of Surrey
Guildford, Surrey GU2 7XH
a.banks@surrey.ac.uk

This study examined the perception of sound quality in different mobile devices (from smart phone to small, then larger tablets) that were matched in physical sound characteristics. A sample of 39 participants were tested on different mobile devices sizes (small- iPhone, medium- iPad Mini and large- iPad) and in different usage contexts (generic content vs. musical training app contexts). Preliminary results showed that the users' perception of sound was affected by device and usage context. However, these differences in perceived sound quality may not be the most important quality to consider within these devices. Instead, considerations such as ease of use seemed to drive considerations for uptake of applications.

Keywords: Mobile devices; sound quality; media-rich design

1. INTRODUCTION

This study investigates research on the impact of using different kinds of devices (smartphones, smaller and large tablets) on user perceptions of sound quality. Previous research (Joseph & Uther, 2009; Uther, 2002; Uther & Banks, 2016; Uther, Zipitria, Uther, & Singh, 2005) had hypothesised that mobile phones may have a natural 'affordance' to listening and speaking functions and hence lead the user to feel a natural bias towards positive impressions of usage of phones to deliver audio-based content. The definitions and measures of affordances were based on that by (Hartson, 2003).

A recent study (Uther & Banks, 2016) explored this with the use of different language learning applications (with differing picture content) and this study sought to extend that research with a study of music-based applications which are also rich in audio content. Based on previous research (Uther & Banks, 2016), it was hypothesised that the perception of sound quality for generic content would be more dependent on device type and sound quality within applications would be more dependent on application type. As physical characteristics were controlled across devices and conditions, any perceived sound differences would be purely subjective and contextually driven.

2. METHOD

2.1 Participants

Thirty nine participants (36 female, 3 males, mean age=20 years,9 non-musicians, 12 novice musicians and 17 expert musicians) were recruited from the student participant pool within the School of Psychology at University of Surrey. The participants were given informed consent and were given tokens for participation which counted towards a lab token scheme that allows the students in turn to recruit participants for their own research in their final year.

2.2 Stimuli and materials

The study tested participants on tasks using an iPhone version 6s, an iPad mini 4 and an iPad Air2. All devices were also loaded with the Auralia Pitch Trainer App (a pitch trainer) and the Clapping App (a rhythm trainer) produced by Reich and Queen Mary University.

A pair of Bose bluetooth headphones (NB: these were not subject to any coding/re-encoding in/out of lossy format) was used to standardise the sound output and sound level was kept constant by taking measurements using a sound level meter. Several questionnaires were used. The first was a general demographic and musical training background questionnaire gathering data about age, gender, mobile device ownership and musical training background. There were bespoke questionnaires

Proceedings of….

(used by Uther & Banks, 2016) on sensory and cognitive affordances that rated the users' perception of sound and picture quality on a 7-point likert scale from the best to worst.

2.3 Design

The study was run as a 'mixed' design, with musical training as group factor (n=9 for those musically untrained and n=27 for participants with some musical training) and device type (iPad vs. iPhone), media type (audio book, music and video) or app type (pitch trainer vs rhythm trainer) as within subjects factors.

2.4 Procedure

Participants were firstly given an informed consent and participant information sheet before proceeding. Data on age demographic, gender, music training background and mobile phone usage were also collected prior to the study commencing using another questionnaire.

Participants then completed a set of tasks on each device (iPhone, iPad mini and iPad) and the order of presentation of each task on each device was randomised. As an index of sensory affordances (Uther & Banks, 2016), participants first rated sound and video quality of generic content on each device. For sound quality, two audio only samples were played: 1) An audio-book sample: a standardised, short, 15 second passage from an audio book ('No.1 Lady's Detective Agency). 2) A musical sample: short, 15 second sample of music from Yo-Yo Ma's rendition of Bach's Cello Suite #1 in G. The audio portion of a video sample was also rated for sound quality. The participant's task was to rate the quality of the sound on the device being played on a likert scale. To rate video quality, a short, 15-second sample of a high-definition video (National Geographic's documentary on the 'Secrets of Antarctica') was played. The participants were asked to rate their perceived video quality on a likert scale. The order of sample and device types tested was randomised across participants.

The participants were then asked to use one of two mobile music training applications for a few minutes and were then asked to rate the software and device suitability. Participants were also asked to explicitly rate the audio and video quality as well as the perceived suitability of each device being used for each software application and the likelihood that they would use that application.

3. RESULTS

3.1 Perception of sound quality in generic content

The subjective comparison of audio and picture quality of generic samples across iPad, iPad Mini, and iPhone was taken as an index of sensory affordances (Uther & Banks, 2016) for each device. The results are presented below. For audio quality, three types of sound were rated for user-perceived quality (audio book, music sample or audio portion of the video clip). Results showed there was a main effect of device on audio rating, with the iPhone having higher rating compared to the iPad mini or iPad ($F_{(1,35)}$ = 5.889, , $p < .05$;see Figure 1 below).

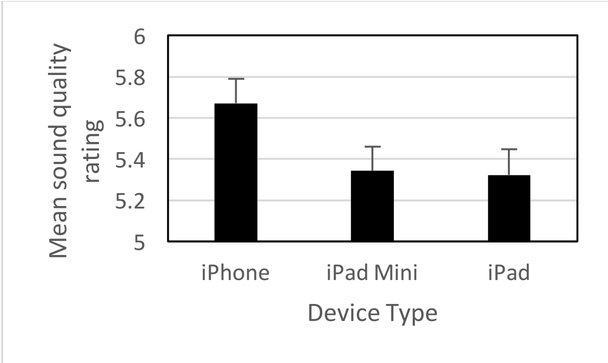

Figure 1: Mean sound quality rating of generic content on each device type.

There was also an effect of media type on sound quality. Despite overall acoustical qualities being controlled, participants rated the audio book sample better than the musical sample and the sound of the video sample as having the worst quality ($F_{(1,35)}$ =28.543, $p < .05$, see Figure 2 below).

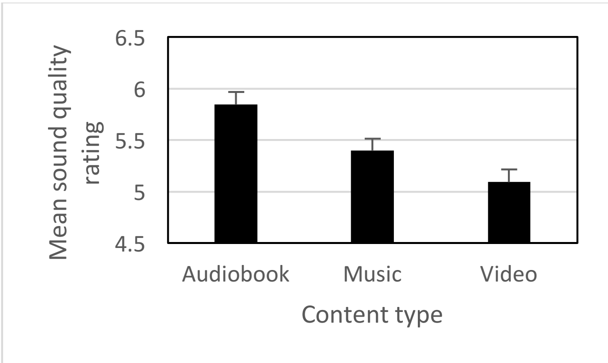

Figure 2: Mean sound quality rating of different kinds of content across all devices.

There was no effect of musical training on the perception of sound quality and no other main effects or interactions were significant.

3.2 Perception of sound quality in training apps

The subjective comparison of audio and video quality of each music training app across the different devices were taken as an index of cognitive affordances (Uther & Banks, 2016) for each device.

The results are presented below. Results showed that quality ratings did not differ between the three devices, suggesting that all types of device afforded equally well for the listening experience within these apps.

There was a main effect of app type on the rating for sound quality, with the Auralia pitch training app having higher audio ratings across all devices compared to the 'Clapping' rhythm app ($F_{(1,36)}$ =26.984, , $p < .01$; see Figure 3).

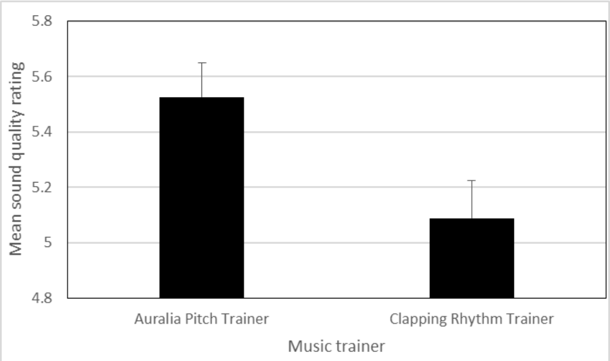

Figure 3: Mean sound quality rating of different kinds of music trainer.

Picture quality was also rate significantly better for the Auralia pitch training app compared to the Clapping app ($F_{(1,36)}$ =11.019, , $p < .01$; see Figure 4).

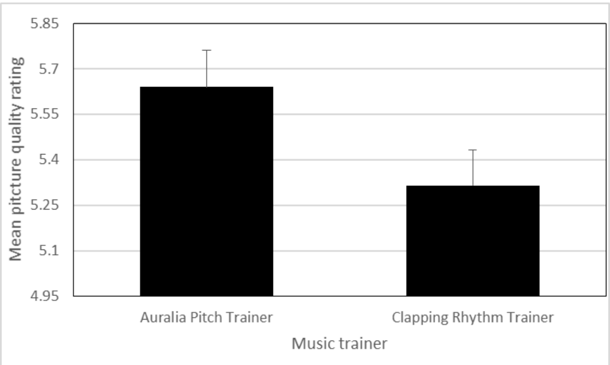

Figure 4: Mean picture quality rating of different kinds of music trainer.

Likelihood of future use was also significantly higher for the Auralia pitch training app compared to the Clapping rhythm app ($F_{(1,36)}$ =8.667, , $p < .01$; see Figure 5).

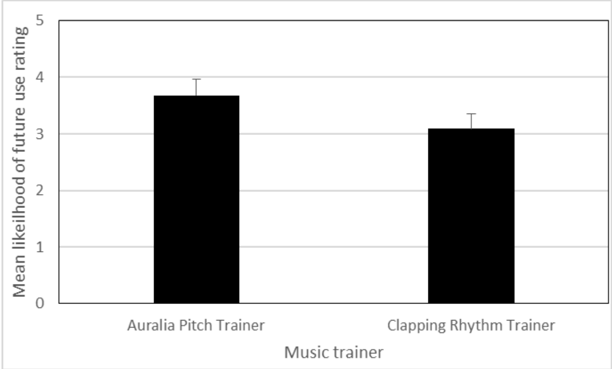

Figure 5: Mean likelihood of future use rating for different kinds of music trainer.

No other interactions or main effects were significant.

4. DISCUSSION AND CONCLUSION

Preliminary results suggest that device type drives the perception of sound quality (even when physical properties are controlled) when listening to generic content. This accords with previous research (Uther & Banks, 2016) in a general sense, although the *direction* of the effect appears to differ. The differences between these and previous data may be attributed to the fact that tablets are more ubiquitous than they used to be and to the fact that the phones have developed further in functionality (label 'smartphone' is much more used) than almost 4 years ago when a similar study was run. Having said that, it is also true that there were individual differences in the data with some individuals showing an opposite trend (such that iPad was perceived as best and iPhone the worst). We have collected some further data on individual trends sound quality as a function of device size and this will be analysed with further categorical analysis to determine what proportion of participants have either type of trend. It is interesting to observe that participants rarely equally rated all devices as the same.

In terms of the data on application type, it was clear that this also mirrored the data we previously collected with language learning apps (Uther & Banks, 2016) such that participants did not rate the sound quality overall as being different but instead their ratings appeared to be more driven by application type. In this study (unlike our previous work), neither application was particularly picture-rich. However, from users' subjective comments at the time of testing, they often commented that they found the clapping app difficult compared to the other application. This could well be likely to clouding their judgement of sound (and even picture) ratings.

What is clear from this data is that purely physical qualities of a stimuli do not drive the rating of sound quality as one might initially expect. Instead, it appears that the perception of sound is influenced by physical context (device type) and also by user impression of the software (whether the software was more positively received or not). This accords with other data collected within the context of developing car hi-fi systems for example, which showed that users perceived within car samples differently to samples played in a laboratory setting, even when the physical attributes were controlled (Beresford, Ford, Rumsey, & Zielinski, 2006). Similar findings (Toole & Olive, 1994) have also showed that the perception of sound quality of speakers is affected by size and model, when actual physical quality does not differ.

Future work will focus on analysing individual differences in more detail as well as also including more musically untrained participants (n=9 in this sample) to determine whether the preliminary findings here hold more broadly. It may also be interesting to investigate perceived sound quality differences in terms of parameters (other than device or application type) that could mitigated for during the design process such as clarity of signal, temporality/synchronisation, etc. Nonetheless, the findings so far do show promise in informing the extent to which contextual factors might affect sound quality judgments, which is useful information for informing the design of audio-rich applications.

5. REFERENCES

Beresford, K., Ford, N., Rumsey, F., & Zielinski, S. (2006). Contextual effects on sound quality judgements: listening room and automotive environments. In *Audio Engineering Society 120th Convention*. Paris, France: Audio Engineering Society.

Hartson, R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour Information Technology*, *22*(5), 315–338. https://doi.org/10.1080/01449290310001592587

Joseph, S. R. H., & Uther, M. (2009). Mobile devices for language learning: multimedia approaches. *Research and Practice in Technology Enhanced Learning*, *4*(1), 7–32. https://doi.org/10.1142/S179320680900060X

Toole, F. E., & Olive, S. E. (1994). Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests, and Other Interesting Things. In *Audio Engineering Society 97th Convention*. San Francisco: AES.

Uther, M. (2002). Mobile Internet usability: what can “mobile learning” learn from the past? *Proceedings IEEE International Workshop on Wireless and Mobile Technologies in Education*, 174–176. https://doi.org/10.1109/WMTE.2002.1039247

Uther, M., & Banks, A. P. (2016). The influence of affordances on user preferences for multimedia language learning applications. *Behaviour & Information Technology*, *35*(4), 277–289. https://doi.org/10.1080/0144929X.2016.1151077

Uther, M., Zipitria, I., Uther, J., & Singh, P. (2005). Mobile adaptive CALL (MAC): a case-study in developing a mobile learning application for speech/audio language training. In *Wireless and Mobile Technologies in Education 2005 WMTE 2005 IEEE International Workshop on* (p. 5 pp.). Ieee. https://doi.org/10.1109/WMTE.2005.46