


Article

The Role of Subjective Quality Judgements in User Preferences for Mobile Learning Apps

Maria Uther ^{1,*} and Sari Ylinen ^{2,3} ¹ Department of Psychology, University of Wolverhampton, Wolverhampton WV1 1LY, UK² Faculty of Educational Sciences, University of Helsinki, P.O. Box 9, 00014 Helsinki, Finland; sari.ylinen@helsinki.fi³ Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, P.O. Box 63, 00014 Helsinki, Finland

* Correspondence: m.uther@wlv.ac.uk

Received: 20 November 2018; Accepted: 18 December 2018; Published: 24 December 2018



Abstract: This study investigated whether subjective quality judgements on sound and picture quality across three devices (iPhone, iPad, and iPad mini) affected user preferences for learning applications. We tested 20 native Finnish-speaking users trialing generic audio clips, video clips, and two kinds of learning apps that were heavily reliant on sound. It was found that there was a main effect of the device on perceived sound quality, replicating earlier findings. However, these judgements did not impact on the users' preferences for different devices nor on their preferences for different applications. The results are interpreted as indicating that perceived quality and affordances are less important for users in these contexts than other considerations (e.g., convenience, mobility, etc.).

Keywords: mobile learning; audio interaction; subjective perception

1. Introduction

The field of 'mobile learning' initially grew from the field of e-learning, as applied to the use of mobile phones to develop applications for learning, rather than to that of non-portable personal computers (PC) (note: the term 'mobile learning' is not used to denote a special kind of learning, but rather to describe learning that is supported by use of mobile devices). Mobile phones were thought to be advantageous for learning in several ways, e.g., portability, mobility, potential for location-based services, etc. (see [1–8] for a review). However, there were other ways in which mobile phones were deemed to offer advantages over fixed PC-based learning. For example, Uther [8] highlighted the potential of mobile phones in terms of 'affordances' for audio interaction. For example, as a phone is built for speaking and listening it may lend itself to superior implementation of audio-based applications. Since those early days, there has been further development of technology such that the term 'mobile learning' now encompasses a myriad of small, portable devices (e.g., small and large tablets and, more recently, 'phablet'-type devices). The boundary between 'phone' and mobile devices has become far more blurred.

The exploration of learner experiences with different mobile devices (phones versus tablets) is of interest as it is clear that the use of handheld devices (especially mobiles) for learning is on the rise [1]. On the other hand, mobile phones have appeared to educators less suited to pedagogical use compared to larger screen devices [9]. This is of course likely due to the fact that perception of what suits 'learning' from a teacher's point of view is necessarily going to entail visual content, whereas, for some fields (e.g., language or musical learning), the reliance on visual content is not as important, and the natural 'affordance' of the necessary device should be anchored more towards the device's

capability in terms of audio content delivery. The concept of affordance is critical to the learning outcomes and is discussed in more detail in references [10,11] for example.

This study builds on previous studies [10,12] that explored affordance of phones versus tablets for learning applications that heavily used sound. Those studies highlighted differences in user perceptions of sound quality, which may in turn impact on the learning and user experience. The first study by Uther and Banks [10] found that sound quality was perceived to be different for different device types (smartphone versus tablet) even when physical acoustical qualities were controlled for. At that time, generic audio clips were perceived as being inferior in quality when delivered on a phone as compared to a tablet. On the other hand, research done a few years later [12] showed somewhat different trends, with phones rating higher in perceived sound quality compared to tablets. The discrepancy between the different studies was attributed to the different positioning in terms of markets for phones and smartphones in the intervening years (smartphones were much more ubiquitous in 2017 than in 2013 when testing was done). The earlier study investigated the perceived sound quality of different English-language learning apps and, although they examined first-language (L1) and second-language (L2) speakers of English, that study used L2 speakers that were already quite experienced in English as they were living in the UK and used the language on a daily basis.

This study sought to extend the previous studies by investigating the importance of perceived sound quality judgements for mobile learning applications that use sound in a non-UK sample. Unlike the previous studies, this study contrasted perceptions of both music and language learning applications within the same study and also used a more naïve group of L2 speakers of English (those living in Finland). This latter manipulation was important as we felt it was important to determine whether sound quality judgements of words were determined by language background. In a previous study with children [13] using a speech-training application [14], it was found that non-native speech samples were related more poorly. Hence, the following hypotheses were formulated in relation to this study:

H1: That Finnish speakers might rate sound quality of the English language content more poorly compared to any Finnish language content.

H2: That there would be a difference in perception of sound quality between phone and tablet categories of devices.

H3: That the sound quality might affect users' rating of the likelihood of future use as well as suitability of the device to the application.

2. Materials and Methods

2.1. Participants

Twenty participants (15 female, 5 males, mean age 30 years, age range between 22 and 53 years) were recruited from the University of Helsinki via email distribution lists and social media. The sample consisted primarily of employed individuals and included seven graduate students. Out of all participants, only nine owned an iPhone (the rest owned other brands of smartphones). We did not have to exclude any participants as a result of diagnosed or possibly undiagnosed hearing impairments in our screening questionnaire. One participant out of 20 owned an iPad mini and five participants out of 20 owned an iPad, with one participant owning an Android tablet. In accordance with ethical guidelines, the participants gave written informed consent and were given a free movie ticket worth approximately €10 for their time. They were free to withdraw at any time without penalty.

2.2. Stimuli and materials

The study used mobile devices which had identical sound output capabilities and similar user interface and interaction styles. These were: an iPhone version 6s, an iPad mini 4, and an iPad Air 2.

The mobile devices were also loaded with the Learn English App (produced by the British Council) and the Auralia Pitch Trainer App.

In order to keep the sound output constant across devices, a pair of Bose Bluetooth headphones was used. In addition, a sound level meter measured the peak, mean and range of the generic sound samples played (which were normalized in advance using audio editing software). This was done to calibrate the ranges of the stimuli to an approximately equal level using the recorded decibel levels.

We used several questionnaires derived from [10,12]. There were general demographic, language, and musical training background questionnaires. These questionnaires gathered data about age, gender, mobile device ownership, language background, and musical training background. There were also questionnaires from the work of Uther & Banks [10] on sensory and cognitive affordances, which rated the users' perception of sound and picture quality on a seven-point scale rating from best to worst. These scales for sound quality were developed with the International Telecommunication Union standards ITU-R BS.1284-1 [15] in mind. Those standards recommend listener grading scales from 'Bad' to 'Excellent.' A7, rather than a five-point scale of rating, was also chosen, as it has been shown to better map the configural space of participant perceptions [16].

2.3. Design

The study was run as a 'mixed' design, with musical training as group factor. Using the questionnaire data, the participants were categorized as: (1) non-musicians, (2) amateur musicians with no formal training but some occasional musical practice, and (3) musicians with formal musical training (e.g., passed accredited music exams). The subject factors included device type (iPad, iPad mini, or iPhone), media type (music, audiobook, or video), and training type (music versus language training app).

2.4. Procedure

The participants were given an informed consent and a participant information sheet before the session started. They were then given the demographic, language, and music background questionnaires. The participants then completed a set of pre-defined tasks on each device (iPhone, iPad mini, and iPad), with the order of presentation of each task on each device randomized. They first rated generic sound and video quality on each device (as an index of sensory affordance). To test sound quality, two audio samples were played: (1) A standardized 15 second passage from the audio-book 'No.1 Lady's Detective Agency' (2) A standardized 15 second musical sample from Yo-Yo Ma's rendition of Bach's Cello Suite #1 in G. They were also given two word-lists of the same words in their native Finnish language and non-native English to rate for sound quality. To rate video quality, a short, 15 second sample of a high-definition video of the same audio musical piece (Bach Cello Suite #1 in G) being performed was played. The participants' headphones were kept in a constant position throughout the testing session and across all devices. The participants filled in a questionnaire to give ratings as they heard each sound.

The participants were then asked to use one of two mobile learning applications for a few minutes. Following this, they were asked to 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 on a seven-point scale, as in references [10,12]. Finally, the participants were also asked to rank each device (iPhone, iPad mini, and iPad) according to which they felt would give the best sound quality. When asked this question, they were allowed to give 'same' rankings to two or three devices at once. The whole experiment took no more than 30 min.

3. Results

We first compared the subjective ratings of audio and picture quality of generic samples across iPad, iPad mini, and iPhone as an index of sensory affordances [10]. For audio quality, three types of sound were rated for user-perceived quality (audio book, music sample, or audio portion of the video

clip playing the same piece of music as the audio clip). The data were analyzed with repeated measures ANOVA using device type and stimulus type as within-subject factors. We also analyzed the group factor of musical proficiency. The analyses for generic content quality rating were performed separately from the application rating, and Bonferroni corrections for family-wise error rates were applied for pairwise multiple comparisons.

There was only a main effect of device type on audio rating, with the iPhone having a lower sound quality rating compared to the iPad mini or iPad ($F_{(1,19)} = 6.333, p < 0.05$; see Figure 1 below). Post-hoc Bonferroni-corrected pairwise comparisons showed that there was no statistically significant difference between the iPad mini and iPad devices.

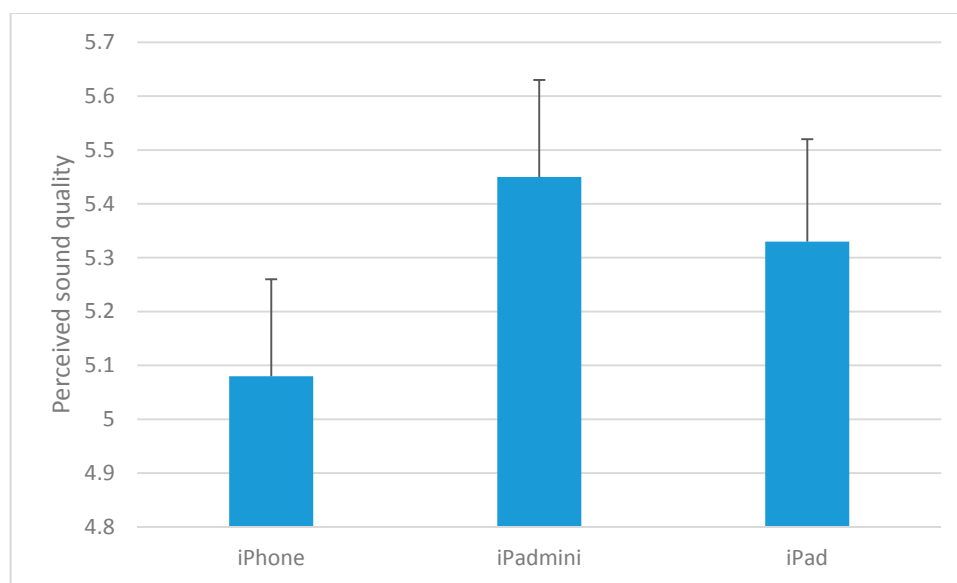


Figure 1. Mean perceived sound quality for generic content as a function of device type.

Further multivariate analyses of other measures did not show any statistically significant different results between different application types (musical training versus language training) nor device type interactions. In general, the applications were rated relatively high (mean rating between 5.5 and 5.0 on a Likert scale of 1–7), but by contrast, the likelihood of future use ratings was much more modest (mean between 3.6 and 4.0 on a Likert scale of 1–7). The latter modest ratings likely reflect the fact that there was little use for the applications, especially—as most participants noticed—because they found the applications too basic for their needs. There were also no significant differences in perceived sound quality between different language stimuli (native versus non-native).

4. Discussion

In this study, we addressed three hypotheses. Our first hypothesis was that Finnish speakers might rate the sound quality of English language speech more poorly compared to native language speech. This hypothesis was not supported. We were surprised as this contrasted with data from another study using a different language learning application with children who were Finnish speakers of English [13]. The contrast in findings between the two Finnish studies could be due to the fact that this sample was already fairly proficient compared to the child sample we tested in the other study. Even though the stimuli were the same, it was clear that the adults in this study had language proficiency that was better than that of the children in the other study. Indeed, the participants here had universally learned English from about age nine to adulthood, so there were very clear differences in the level of English proficiency. In order to see similar patterns as we did with the child sample, we will likely need to test adult participants with a lower level of proficiency to match the child sample. Another possibility is that children cannot tell the difference between difficulties caused by sound quality and

those due to their own (poor) non-native language skills, whereas the adults might better realize that their difficulties are due to their skills and not to sound quality. In this sense, adults may be more objective in their sound quality judgements.

With respect to the second hypothesis regarding the perceived sound quality using different devices, the pattern of data here was somewhat different from that found in the most recent study by our lab [12] with a UK-based sample. On the other hand, the pattern of results was more similar to that found in the first study [10]. These differences could be because the market penetration of Apple devices is different for the United Kingdom compared to Finland. One analysis showed that UK mobile users used iOS more than Android, whereas the reverse was true for Finland [17]. This was also supported by the fact that there were very few iPad users within the sample ($n = 5$) and also fewer iPhone users ($n = 8$). Although device brand did not appear to affect the ratings, the small sample of iOS tablet users might not be large enough to reflect any real effects. Still, the most likely explanation for the discrepancy with more recent findings is that iPads were rated primarily by non-users—and so they might have been perceived as ‘better’ if, for example, they are perceived as superior products (e.g., because of their price, which might be out of the users’ reach). It is important to note that the mobile devices compared here differed in price and market position, which might be the most likely cause of the perceived differences. This is not unlike the effects seen in other contexts of sound quality tests, where branding appeared to affect users’ ratings [18,19].

Finally, the last hypothesis that sound quality might affect the users’ impressions of suitability and future use was not supported. The data clearly showed that despite perceived differences, this did not appear to affect the learners’ perception of suitability of a learning application for a particular device nor their desire to use such an application in the future. Any ratings on attributes such as suitability or sound quality did not impact on the likelihood of using the learning application. Taken in the context of the field of mobile learning, this is an interesting finding as it suggests that learners may opt for convenience of use as a primary criterion for the selection of an application [20]. Despite the fact that the implementation of mobile learning applications is not without challenges or even negatives from a learner [21,22] or even teacher perspective [9], mobile devices have an appeal for affording portability, collaboration, and ‘just-in-time’ interaction. In terms of future research, although it is important to examine device and learner interactions as we have done here, we also need to examine the extent to which mobile learning is adopted in contexts characterized by specific socio-cultural factors which have been argued to be critical for the successful implementation of mobile learning applications [23].

5. Conclusions

In summary, it appears that sound quality is not tied to purely physical qualities, as was found in other studies. Moreover, users perceive sound differences between devices even when none exists. In terms of the central focus of this study regarding mobile learning, the extent to which subjective differences in sound-quality experience impact on the actual learning outcomes is not directly determined by this study. In all likelihood, the factors examined do not appear to be driving consumer choices and overall impressions of suitability. Future research may, for example, use a different approach (e.g., using qualitative measures) to examine in more detail the extent to which differences in subjective experiences are affected by aspects such as branding, previous experience, etc. More importantly, determining the order of priority of features for the user (e.g., portability, affordance for collaboration, etc.) is also critical for the successful planning of mobile learning applications. However, only further studies with larger samples and tied to actual learning outcomes will be capable of determining the impact of different factors on educational outcomes.

Author Contributions: Conceptualization, M.U. and S.Y., methodology, M.U., S.Y., statistical analysis, M.U., writing, original draft preparation, M.U., writing—review & editing, S.Y.

Funding: This research was funded by the Nokia Foundation Visiting Professor Grant awarded to Professor Uther for a visit to Cicero Learning, University of Helsinki, Finland in 2018.

Acknowledgments: The authors acknowledge the assistance of Mari Tervaniemi in practical matters related to accommodating and hosting Uther’s research visit to Finland. The authors also acknowledge the assistance of Anna-Riikka Smolander of the University of Helsinki in relation to participant recruitment.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Wu, W.H.; Wu, Y.C.J.; Chen, C.Y.; Kao, H.Y.; Lin, C.H.; Huang, S.H. Review of trends from mobile learning studies: A meta-analysis. *Comput. Educ.* **2012**, *59*, 817–827. [[CrossRef](#)]
2. Kukulska-Hulme, A. Mobile usability in educational contexts: What have we learnt? *Int. Rev. Res. Open Distance Learn.* **2007**, *8*, 1–16. [[CrossRef](#)]
3. Orr, G. A review of literature in mobile learning: Affordances and constraints. In Proceedings of the 6th IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education, Kaohsiung, Taiwan, 12–16 April 2010; pp. 107–111.
4. Jacob, S.M.; Issac, B. The Mobile Devices and its Mobile Learning Usage Analysis. In Proceedings of the International Multiconference of Engineers and Computer Scientists, Hong Kong, China, 19–21 March 2008; Volume I.
5. Sung, Y.T.; Chang, K.E.; Liu, T.C. The effects of integrating mobile devices with teaching and learning on students’ learning performance: A meta-analysis and research synthesis. *Comput. Educ.* **2016**, *94*, 252–275. [[CrossRef](#)]
6. Sharples, M.; Pea, R. Mobile learning. In *The Cambridge Handbook of the Learning Sciences*, 2nd ed.; Sawyer, R.K., Ed.; Cambridge University Press: New York, NY, USA, 2015; pp. 501–521.
7. Koole, M.; Buck, R.; Anderson, K.; Laj, D. A Comparison of the Uptake of Two Research Models in Mobile Learning: The FRAME Model and the 3-Level Evaluation Framework. *Educ. Sci.* **2018**, *8*, 114. [[CrossRef](#)]
8. Uther, M. Mobile Internet usability: What can ‘mobile learning’ learn from the past? In *Proceedings—IEEE International Workshop on Wireless and Mobile Technologies in Education*; IEEE Computer Society: Washington, DC, USA, 2002; pp. 174–176.
9. Şad, S.N.; Göktaş, Ö. Preservice teachers’ perceptions about using mobile phones and laptops in education as mobile learning tools. *Br. J. Educ. Technol.* **2014**, *45*, 606–618. [[CrossRef](#)]
10. Uther, M.; Banks, A.P. The influence of affordances on user preferences for multimedia language learning applications. *Behav. Inf. Technol.* **2016**, *35*, 277–289. [[CrossRef](#)]
11. Hartson, R. Cognitive, physical, sensory, and functional affordances in interaction design. *Behav. Inf. Technol.* **2003**, *22*, 315–338. [[CrossRef](#)]
12. Uther, M.; Banks, A.P. User Perceptions of Sound Quality: Implications for the Design and Use of Audio-Based Mobile Applications. *Int. J. Hum. Comput. Interact.* **2018**. [[CrossRef](#)]
13. Uther, M.; Smolander, A.-R.; Junntila, K.; Kurimo, M.; Karhila, R.; Enarvi, S.; Ylinen, S. User Experiences from L2 Children Using a Speech Learning Application: Implications for Developing Speech Training Applications for Children. *Adv. Hum.-Comput. Interact.* **2018**, *2018*, 7345397.
14. Karhila, R.; Ylinen, S.; Enarvi, S.; Palomäki, K.; Nikulin, A.; Rantula, O.; Viitanen, V.; Dhinakaran, K.; Smolander, A.-R.; Kallio, H.; et al. SIAK—A Game for Foreign Language Pronunciation Learning. In *Interspeech 2017*; ISCA: Stockholm, Sweden, 2017.
15. International Telecommunications Union. *General Methods for the Subjective Assessment of Sound Quality*; International Telecommunications Union: Geneva, Switzerland, 2003.
16. Maxell, M.S.; Jacoby, J. Is there an optimal number of alternatives for likert-scale items? Effects of testing time and scale properties. *J. Appl. Psychol.* **1972**, *56*, 506–509.
17. DeviceAtlas. Android v iOS Market Share 2018. 2018. Available online: <https://deviceatlas.com/blog/android-v-ios-market-share> (accessed on 21 December 2018).
18. Toole, F.E.; Olive, S.E. Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests, and Other Interesting Things. In *Audio Engineering Society 97th Convention*; AES: San Francisco, CA, USA, 1994.
19. Beresford, K.; Ford, N.; Rumsey, F.; Zielinski, S. Contextual effects on sound quality judgements: Listening room and automotive environments. In *Audio Engineering Society 120th Convention*; Audio Engineering Society: Paris, France, 2006.

20. Chang, C.C.; Yan, C.F.; Tseng, J.S. Perceived convenience in an extended technology acceptance model: Mobile technology and English learning for college students. *Australas. J. Educ. Technol.* **2012**, *28*, 809–826. [[CrossRef](#)]
21. Heflin, H.; Shewmaker, J.; Nguyen, J. Impact of mobile technology on student attitudes, engagement, and learning. *Comput. Educ.* **2017**, *107*, 91–99. [[CrossRef](#)]
22. Chu, H.C. Potential negative effects of mobile learning on students' learning achievement and cognitive load—a format assessment perspective. *Educ. Technol. Soc.* **2013**, *17*, 332–344. [[CrossRef](#)]
23. Koole, M.L. A model for framing mobile learning. *Mob. Learn. Transform. Deliv. Educ. Train.* **2009**, *1*, 25–47. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).