User perceptions of sound quality: implications for the design and use of audio-based mobile applications

Maria Uther,[1,2*] & Adrian P. Banks[2]

1. Department of Psychology, University of Winchester, Sparkford Rd, Winchester, Hampshire, SO22 4NR, United Kingdom, tel: +44 1962 624892, Maria.Uther@winchester.ac.uk
2. School of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom, tel: +44 1483 68 9435, a.banks@surrey.ac.uk

**Corresponding author*

Running head: User perceptions of sound quality on mobile devices

Keywords: audio quality, user perceptions, mobile applications

Author biographies:

Maria Uther

Maria Uther is a professor of cognitive psychology at University of Winchester. Her research interests include auditory perception, mobile technologies, speech technologies and mobile technologies for learning. She received her Ph.D. in Psychology from the University of New South Wales, Australia in 1999. Contact her at maria.uther@winchester.ac.uk.

Adrian Banks

Adrian Banks is a senior lecturer in psychology at the University of Surrey where he investigates the cognitive mechanisms underpinning how people think, such as how people reason and make decisions, uses this knowledge to improve their thinking. He received his PhD from the University of Surrey in 2002. Contact him at a.banks@surrey.ac.uk.

Abstract

This study sought to investigate the effect that contextual cues (in particular, device type and content type) have on the perception of sound quality. A sample of 49 participants were tested on different mobile devices sizes (small- iPhone, medium- iPad Mini and large- iPad) which had identical sound output characteristics within in different usage contexts (generic content vs. musical training app contexts). Results showed that the users' perception of generic sound types was affected by device type, with iPhones appearing to have better sound quality compared to larger devices. On the other hand, within application contexts, the application type seemed to affect user perceptions more, with the rhythm training application rating poorer on sound quality, picture quality and likelihood of future use as compared to the pitch training application (although this may be due to the perceived increased difficulty). Together, these findings demonstrate the influence of device and content cues (when actual physical qualities are controlled) on user sound perception. Interestingly, differences in perceived sound quality was not accompanied by an overriding preference for that device as compared to other devices. Instead, considerations such as ease of use seemed to drive considerations for uptake of applications.

1. Introduction

This study investigates research on the impact of using different categories of devices (smartphones, smaller and large tablets) on user perceptions of sound quality. Previous research (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. This concept of an 'affordance' stems from an ecological (Gibson, 1979) characterisation of objects in the visual scene that is interpreted from the perspective of an agent performing an action (e.g. a telephone would afford listening to for people but not for a fish). In this way, the modern mobile phone may have a natural affordance for listening to audio content and speaking into.

At its core, the concept of affordances seems intuitive. However, there are further developments to the original Gibsonian view that are useful within the context of usability and technology. For example, Norman (1988) suggests that although affordances can be apparent visually, there are other functions of an object that aren't immediately cued by the object and yet apparent (e.g. a chair can obviously be sat on but also be stood on, as many use this object in that way through experience). In this way, the Gibsonian view is that affordances cannot be learnt and yet within Norman's view, affordances can be learned (McGrenere & Ho, 2000). The influence of context [1]on the perception of an object lending itself for a particular action is an important one for the use of modern technology, where the perception of an object (e.g. mobile phone) has changed considerably over time. At the outset, the early mobile phones were very much just for listening and speaking to others. These days, modern mobile phones (or smartphones as they are called) are complex, multimedia systems by which we can listen to music, play videos, surf the Internet and so

[1] By context, we refer to emergent properties of an object rather than situational context (i.e. changes that occur in the environment whether when an object is used).

forth. In this way, consideration of the affordances of smartphones is likely to be fairly dynamic and dependent upon situational factors.

A useful framework for understanding the affordance of modern technology comes from the work of (Hartson, 2003) which distinguishes between sensory, cognitive, and educational affordances. Simply put, the educational benefit is dependent upon perceptions at a cognitive and sensory level. This framework of affordances was used by (Uther & Banks, 2016) which studied user perceptions of sensory qualities and how that information is processed within language learning applications. The starting point was subjective sensory and cognitive perceptions rather than specifically being on learning outcomes (c.f. Dickey's, (2003) concept of 'pedagogical affordances' and Hartston's (2003) 'educational affordances'), as the learning outcomes are undoubtedly affected by user perception and uptake.

Previous work (Uther & Banks, 2016) explored the influence of affordances within two different mobile language learning applications (one heavier on audio, and light on pictures, the other more rich in pictures/video). In that study, an iPhone and iPad device was used to compare user perceptions of sound. The contrast of an iPhone and iPad provided an ideal opportunity to compare devices that yielded physical sound quality which was identical, same interaction style and yet differing in device type. That study found that iPhones had lower sound quality experiences for generic content (contrary to the expected hypothesis that phones might be judged to be 'better' for sound). This finding was interpreted as possibly being due to the phone being judged as less 'powerful' compared to the iPad, cueing an expectation of a different kind of affordance. It was also found that the extent to which video content was present affected the perception of the sound. Those findings were also in accordance with previous data from different settings (e.g. hi-fi systems being judged differently according to size or brand, see Toole & Olive, 1994)

1.1. Rationale and hypotheses

This study seeks to extend previous findings (Uther & Banks, 2016) with a study of music-based applications which are also rich in audio content. The use of musical training applications are important not only from the point of view of training music directly, but to the potential broader benefits of this kind of audio training (e.g. with cochlear implants, see Good et al., 2017). Our earlier findings would suggest 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. Since the publication of the previous study, the iPad mini has also emerged as an intermediary between the iPhone and iPad which is useful as a comparison to determine whether it is simple screen/device size (i.e. 'powerfulness') or category (tablet vs. phone) that drives user experience effects. As in our previous study, physical characteristics were controlled across devices and conditions of that any perceived sound differences would be purely subjective. A quick comparison of the features of all three devices is given in Table 1 below.

Table 1. Comparison of physical constraints on the three Apple devices used in the study (values compared from the manufacturer's technical specifications).

Attribute	iPhone6s	iPad mini4	iPad Air2
Size	4.7 inch (diagonal)	7.9 inch (diagonal)	9.7 inch (diagonal)
Weight (g)	143	298.8	437
Video quality	H.264 video up to 4K	H.264 video up to 4K	H.264 video up to 4K
Audio frequency response	20Hz to 20,000Hz	20Hz to 20,000Hz	20Hz to 20,000Hz
Display density	326 pixels per inch	326 pixels per inch	264 pixels per inch

In consideration of previous findings and our review of theory, the following hypotheses were formed:

Hypothesis 1:

If device size drives sound quality judgements, then different devices would be judged as differing in sound quality for generic content in a graded manner, assuming the judgement is relating to the perceived ‘powerfulness’ of the device.

Hypothesis 2:

That the perception of sound might be affected within application content to be more driven by subjective perceptions of the applications rather than device type (e.g. presence of video, overall affective reactions).

Hypothesis 3:

That user preferences for future use might not be as driven by the subjective (sensory) experiences so much as by the overall affective impressions of the software and devices. In other words, the degree to which subjective (sensory) experiences play a part in consumer choice might be overridden by other factors (e.g. portability of the device).

2. Material and Methods

2.1 Participants

Forty nine participants (44 female, 5 males, mean age=20.5 years, age range between 19 and 33,14 non-musicians, 16 amateur musicians and 19 expert musicians) were recruited from the student participant pool within the School of Psychology at University of XXX and via campus noticeboards from the wider university community. The sample were primarily students (n=45) and there were 4 university (non-academic) staff. Within this sample, 36 participants owned an iPhone (and conversely 13 did not). We screened out participants who had diagnosed hearing impairments and did an informal test of the calibrated hearing levels of sounds standardised to about 65dB to ensure that no participant felt these were experientially too loud or soft (and therefore might have indicated an undiagnosed disorder). We did not have to exclude any participants as a result of diagnosed or possibly undiagnosed hearing impairments. Only one participant owned an iPad mini and 6 out of 49 participants owned an iPad. The participants were given informed consent and were given either a £5 Amazon voucher or 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 of which had identical sound output capabilities and similar interface/interaction styles. All devices were also loaded with the Auralia Pitch Trainer App and the Clapping App produced by Reich and Queen Mary University.

A pair of Bose Bluetooth (over ear) headphones was used to standardise the sound output and sound level was kept constant by taking measurements using a sound level meter in a quiet environment. The sound level meter measured the peak, mean and range of the generic sound samples played (which were normalised in advance using audio editing software) using an in-built microphone placed to detect the output of each device from the headphones by placing the microphone between both sides of the headphones to exclude extraneous noise and provide a realistic measurement of the headphone output during normal usage by participants. The ranges of the stimuli were approximately equal, as were the peak and mean decibels.

Several questionnaires were used. The first was a general demographic and musical training background questionnaire (based on the MUSE questionnaire: Chin & Rickard, 2012) gathering data about age, gender, mobile device ownership and musical training background. There were also questionnaires (Uther & Banks, 2016) on sensory and cognitive affordances (that rated the users' perception of sound and picture quality on a 7-point scale rating from best to worst). The scales for sound quality are similar to those recommended by International Telecommunication Union standards ITU-R BS.1284-1 (ITU, 2003) which recommends listener grading scales from 'Bad' to ' Excellent. We also opted for a 7 rather than a 5 point scale of rating as it has been shown to better map configural space of participant perceptions (Maxell & Jacoby, 1972).

2.3 Design

The study was run as a 'mixed' design, with musical training as group factor (participants were coded into one of three groups from their musical background and experience questionnaire: 1) non-musicians, 2) amateur musicians with no formal training but some occasional musical practice or 3) musicians with certification of formal musical training (e.g. passed established music exams). Device type (iPad. iPhone), media type (audio book, music and video) and well as musical training app type (pitch trainer vs rhythm trainer) served 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 (which kinds of devices they owned and how frequently they were used) 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, participants first rated sound and video quality of generic content on each device. For sound quality, two audio 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 #1in G. 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.

Participants kept their headphones on in a constant position throughout each comparison test of the same material across three devices. They were given a clipboard with the questionnaire to fill in their ratings as they heard each sound.

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. 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 on a seven-point scale, as in Uther & Banks (2016). Finally, 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 asking this question, they were allowed to give ‘same’ rankings’. The whole experiment took about 20-25 minutes.

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. 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 were analysed with ANOVA with repeated measures factors of device type, stimulus type and their interactions. We also analysed the group factor of musical proficiency. Analyses were performed separately for generic content quality ratings from the application ratings and Bonferroni corrections for family-wise error rates were applied for pairwise multiple comparisons.

The 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 ($F_{(1,46)}$ =4.447, , $p < .05$;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 important acoustical qualities (intensity, quality of sound output) 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 ($F_{(1,46)}$ =35.140, $p < .05$, 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 or type of device (apple vs. other mobile or tablet) ownership and use on the perception of sound quality. No other interactions or main effects were significant. Despite asking participants how frequently they used their devices, they almost universally used their device several times a day. When looking at ownership/use of iPhones, this did not have an effect on these judgements. We also looked at whether individuals' rankings of what device they believed would give the best sound quality would correspond with their actual ratings on the task and there was no correlation.

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. 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.

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 ($F_{(1,46)}$ =29.321, , $p < .01$; see Figure 3).

Figure 3: Mean sound quality rating of the two types of music trainer.

Picture quality was also rate significantly better for the Auralia pitch training app compared to the Clapping app ($F_{(1,46)}$ =8.574, , $p < .01$; see Figure 4).

Figure 4: Mean picture quality rating of the two types types of music trainer.

Likelihood of future use was also significantly higher for the Auralia pitch training app compared to the Clapping rhythm app ($F_{(1,46)}$ =11.440, , $p < .01$; see Figure 5).

Figure 5: Mean likelihood of future use rating of the two types of music trainer.

No other interactions or main effects were significant and nor did musical training or phone/tablet usage appear to have an effect on these results. Hence, there was no significant preference for a particular device in terms of likelihood of future use.

4 Discussion

These data suggest that differences in device types affect 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 differs to our previous studies such that iPhone was perceived as better in sound quality compared to either tablet devices. The differences between these and previous findings 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 (the label 'smartphone' is much more used) than almost 4 years ago when the Uther & Banks (2016) study was run. Having said that, it is also true that there were individual differences in the data with a small subset (about 5-8 depending on the condition) of individuals showing an opposite trend (such that iPad was perceived as best and iPhone the worst). We also asked participants to explicitly rate their expectations of sound quality of the overall devices (ranking sound quality expectations of iPhone, iPad Mini and iPad on a scale of best to worst and they were given the opportunity to rate all or some equally). Interestingly, their perceptions of overall device quality did not correspond with their actual judgements of generic sound quality. In other words, individuals were poor predictors of their own judgements.

In general terms however the first hypothesis (that generic sound quality judgements are affected by device type) was supported by the results. However, the hypothesis that if differences in device type were related to directly to size (and therefore perceived 'powerfulness') was not borne out by the data. Instead, back to the original hypothesis from our 2016 paper, it would appear that smartphones were perceived at delivering 'superior' sound quality. As the two tablets (iPad mini and iPad were rated similarly), this is likely to be related to the perceived affordance of the smartphone as the device most commonly used for listening to music and talking phone calls. Naturally weight and size has to differ between these smartphone and tablet categories, but it would be impossible to devise an experiment that contrasted a smartphone that was matched to a tablet in size and weight. Weight and size are necessarily part of the features that distinguish between these device categories.

In terms of the data on sound quality within the music training applications, it was clear that this also mirrored the data we previously collected with language learning apps in 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 such that the rhythm trainer was perceived as being reduced in sound quality compared to the pitch trainer. Hence, the second hypothesis (i.e. that user impressions of software drive sound quality judgements more than device type) is supported.

One may wonder why sound quality appears to be rated differently within and outside training applications. It should be noted that for generic sound ratings, the users are not actively interacting with the interface on the device at all (indeed, within this paradigm, the experimenter is playing the stimuli to the participant and they do not see the screen at all). On the other hand, within an application, the participant is actively interacting with the interface and have total control of the device and therefore be more susceptible to contextual factors such as the presence of visual stimuli or affective reactions to the software.

In previous work, perceived differences in sound quality within applications were attributed to the differences in picture/video content. However, in the present study, neither of the two applications tested were particularly picture-rich, so this explanation is not the most likely. Instead, the differences may be due to affective reactions to each software. It was clear from users' subjective comments at the time of testing, they often commented that they found the rhythm training app more difficult compared to the other application. The perceived increase in difficulty for the rhythm application might then in turn result in reduced positive affect and consequent perception of poorer sound quality. Another possibility is also that the kind of sounds used within applications matters. The rhythm training app had less melodic stimuli compared to the pitch trainer application. This could also in turn result in differences in the sense of melodic stimuli perhaps sounding 'clearer' than claps. However, it is interesting to note that picture quality was also rated better in the Auralia Training App, so that finding would lend weight to the idea that it is generic affective responses driving higher sound quality ratings. Further testing (e.g. with different kinds of musical stimuli) would be necessary to directly test whether the effects are likely to be due to affective responses or instead driven by the heavy use of pitch stimuli within the Auralia application. Another (albeit remote) possibility is that manufacturer specifications on technical specifications may not be as uniform as thought, so independent verification of manufacturer specifications would be useful to exclude this possibility.

In terms of how affordance ratings affect user perceptions of suitability for each, it appears that we did not see that there was any particular effect of different software being rated differently in suitability for each device, contrary to previous findings. This is likely to be due to the similar level of picture content (both rated and actual) for each software, so that there were no particular preferences for device being suited to any kind of device. In our previous study, the fact that one software was more picture/video rich lent itself to a stronger preference for the larger screen. In the current study, as neither software was especially picture/video rich, none of the devices were rated as more suitable for any of the software. In terms of likelihood of future use, there was a tendency for the pitch trainer to be more rated more likely to be used. As for the sound ratings, this might be most likely related to the more positive affective reactions to this software. In terms of hypothesis 3, it would appear that affective reactions are at least to some extent driving the learner's predicted future use above the considerations of the perceived (generic sensory) qualities.

Hence, it would appear that generic sensory or cognitive affordance measures do not always translate to decisions to prefer a particular mobile device – in this dataset, the sound quality for generic sound quality was judged to be poorer for the tablet devices, but this did not translate to a preference for using those types of devices. Here, other considerations such as convenience, mobility ease of software use appear to override quality concerns. This argument regarding convenience is also supported by views that mobile learning affords ubiquity (anytime, anywhere) over any other physical aspect (Lai, Yang, Chen, Ho, & Chan, 2007; Orr, 2010; Ko, 2017).

Interestingly, all of these subjective perceptions appear not to be affected by musical experience nor even ownership/usage of these kinds of devices as neither of those factors had an effect on the sound quality judgements. For future research, there are two obvious areas of priority: 1) to look at actual physical differences on performance and user satisfaction. Although it wasn't the focus of this study, it has been shown that even real physical differences in audio characteristics might not necessarily affect learner outcomes, the physical differences can impact learner satisfaction (Ritzhaupt, Gomes, & Barron, 2008). 2) look at actual learning outcomes as opposed to just physical/cognitive affordances. In the present study, the focus was not on educational outcomes - although arguably poor user satisfaction would likely affect educational outcomes by in turn impacting on uptake and frequency of usage.

What is clear from this data is that purely physical qualities of stimuli do not drive the rating of sound quality as one might initially expect. Instead, it appears that the perception of generic sound content is influenced by physical context (device type) and also by user impression of the software (and in turn, likely to be a function of 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 by (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. Although branding is the same here for our devices (they are all Apple), perceived differences in pricing (and hence powerfulness/quality) for example, might differ. Hence, the findings support the view that contextual factors affect sound quality judgments (Fiebig, 2015), which is useful information for informing the design of audio-rich applications.

5. Conclusions

These data have clear practical applications for the design of audio-rich applications. It is clear that for audio-rich applications, pure physical characteristics of the sound stimulus are not the only factor in the user's experience of sound. Given that the addition of audio to games has been shown to be important for learner engagement (Byun & Loh, 2015), this is an important consideration for the design of learning applications. The findings show that subjective perception of audio is affected by context and is independent of the physical constraints. It seems that the perception of sound quality is (at least in this sample), affected by perceived expectations (in turn driven probably by affordances) of a device, but this may be the case mainly for audio-only content. Secondly, it would appear that the importance of perceived inferior audio quality might be downplayed in the context of other aspects such as user impression of software or portability considerations for example. These data are useful not only in the design of musical learning applications, but also in other learning applications where audio-based content is important (e.g. speech learning applications).

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.

Byun, J., & Loh, C. S. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers in Human Behavior*, *46*, 129–138. https://doi.org/10.1016/j.chb.2014.12.052

Chin, T., & Rickard, N. S. (2012). The Music USE (MUSE) Questionnaire: An Instrument to Measure Engagement in Music. *Music Perception: An Interdisciplinary Journal*, *4*, 429–446. https://doi.org/10.1525/mp.2012.29.4.429

Dickey, M. D. (2003). Teaching in 3D: Pedagogical Affordances and Constraints of 3D Virtual Worlds for Synchronous Distance Learning Constructivist Learning Environments. *Distance Education*, *24*(1), 105–121.

https://doi.org/10.1080/0158791032000066552

Fiebig, A. (2015). Influence of context effects on sound quality assessments. In *Euronoise* (pp. 2555–2560). Maastricht: Belgian Acoustical Association (ABAV) and Acoustical Society of the Netherlands (NAG). Retrieved from http://www.conforg.fr/euronoise2015/proceedings/data/index.html

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston, USA: Houghton Mifflin.

Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing*, *in press*.

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

International Telecommunications Union. (2003). *General methods for the subjective assessment of sound quality*. Retrieved from https://www.itu.int/rec/R-REC-BS.1284-1-200312-I/en

Ko, M.-H. (2017). Learner perspectives regarding device type in technology-assisted language learning. *Computer Assisted Language Learning*, *30*(8), 844–863. https://doi.org/10.1080/09588221.2017.1367310

Lai, C. H., Yang, J. C., Chen, F. C., Ho, C. W., & Chan, T. W. (2007). Affordances of mobile technologies for experiential learning: The interplay of technology and pedagogical practices. *Journal of Computer Assisted Learning*, *23*(4), 326–337. https://doi.org/10.1111/j.1365-2729.2007.00237.x

Maxell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, *56*(6), 506–509.

McGrenere, J., & Ho, W. (2000). Affordances: Clarifying and evolving a concept. In

Proceedings of Graphics Interface 2000 (pp. 179–186).

Norman, D. (1988). *The Design of Everyday Things*. New York: Basic Books.

Orr, G. (2010). A review of literature in mobile learning: Affordances and constraints. In *6th IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education, WMUTE 2010: Mobile Social Media for Learning and Education in Formal and Informal Settings* (pp. 107–111). Kaohsiung: IEEE. https://doi.org/10.1109/WMUTE.2010.20

Ritzhaupt, A. D., Gomes, N. D., & Barron, A. E. (2008). The effects of time-compressed audio and verbal redundancy on learner performance and satisfaction. *Computers in Human Behavior*, *24*, 2434–2445. https://doi.org/10.1016/j.chb.2008.02.017

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. 191-195). IEEE. https://doi.org/10.1109/WMTE.2005.46