



The influence of affordances on user preferences for multimedia language learning applications

Maria Uther & Adrian P Banks

To cite this article: Maria Uther & Adrian P Banks (2016) The influence of affordances on user preferences for multimedia language learning applications, Behaviour & Information Technology, 35:4, 277-289, DOI: [10.1080/0144929X.2016.1151077](https://doi.org/10.1080/0144929X.2016.1151077)

To link to this article: <http://dx.doi.org/10.1080/0144929X.2016.1151077>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 16 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 244



View related articles [↗](#)



View Crossmark data [↗](#)

The influence of affordances on user preferences for multimedia language learning applications

Maria Uther^{a,b}  and Adrian P Banks^b

^aDepartment of Psychology, University of Winchester, Winchester, UK; ^bDepartment of Psychology, University of Surrey, Surrey, UK

ABSTRACT

This study investigates the influence of sensory and cognitive affordances on the user experience of mobile devices for multimedia language learning applications. A primarily audio-based language learning application – ‘Vowel Trainer’, was chosen against a comparison, text and picture-based language learning application – ‘Learn English for Taxi Drivers’. Impressions of the two applications were assessed on two different devices that have virtually the same interface and identical sound output (when headphones are used), but differ in physical size: the iPhone and the iPad. A mixed design was chosen, with native language as a group factor and device type (iPad vs. iPhone) and language application type (audio vs. video) as within groups factors. Assessments of sensory and cognitive affordances were made, along with measurement of learner preferences of each application. Data from 41 participants (21 native English speakers, 20 non-native English speakers) were analysed, revealing device differences in both audio and visual subjective quality ratings, despite only visual quality being affected by the device’s physical limitations. We suggest that sensory affordances (indexed by subjective quality) are not simply a function of physical limitations, but are heavily influenced by context. The implications for developing design guidelines for language learning and other multimedia applications are discussed.

ARTICLE HISTORY

Received 4 September 2015
Accepted 2 February 2016

KEYWORDS

Perceived sound quality;
mobile language learning;
affordances

1. Introduction

Within the field of language learning, there have been several advancements in mobile technology that have culminated in the development of software to assist learners of English as a second language. As important as these developments are, language learning technology has historically used simple interfaces teaching simple vocabulary and grammar skills. However, important work is now being done to expand the repertoire of language learning applications to incorporate multimedia elements; for example, the ‘Learn English’ application by the British Council and the grammar skills tool by Guerrero, Ochoa, and Collazos (2010). Further use of multimedia elements in language teaching is of course particularly important in teaching spoken language, as it requires heavy audio use.

The range of spoken language learning applications has historically been PC-based and usually specialises in phonetic training (Cook 2008). For example, the SPATS programme (Miller et al. 2008; Watson et al. 2008) is targeted at remedial work with hearing aid users. The work on the popular ‘high variability phonetic

training’ (whereby learners are given targeted training on non-native speech sounds) is also notable in the field. It was initially designed to train phonetic discrimination in Japanese speakers of English, but later has been extended to other language groups as well. High variability phonetic training has proven to be a very successful, theoretically well-grounded and popular technique for speech training (see Bradlow et al. 1997, 1999; Giannakopoulou, Uther, and Ylinen 2013; Iverson, Hazan, and Bannister 2005; Iverson and Evans 2009; Lively, Logan, and Pisoni 1993; Lively et al. 1994; Uther et al. 2007; Ylinen et al. 2010) and is the technique behind one of the mobile device applications studied in this research project (the UCL Vowel Trainer).

Interestingly, although there is an abundance of research on learning outcomes of these newer language learning technologies incorporating multimedia, there has been little research on the *design* and *evaluation* of the usability of such applications. Some design work has been done with simple, visually based applications in mind (Chinnery 2006; Fisher et al. 2009; Kukulska-Hulme 2005; Kukulska-Hulme and Shield 2008; Thornton and Houser 2005), but the generalisability

CONTACT Maria Uther  maria.uther@winchester.ac.uk

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

to audio-based applications has not been directly empirically tested. From an *evaluation* point of view, although there have been several attempts to evaluate computer-assisted language learning applications in general (Hubbard 1987; Plass 1998; Strei 1983), the focus has been mainly on producing checklists for quality, which do not suffice to adequately evaluate the suitability of the device or software for the intended purpose (Oliver 2000). Also, current checklists need updating to incorporate usage for non-standard portable devices, which are now becoming ubiquitous (mobile phones, iPods, iPads, etc.). Given the suggestion that some of these devices might actually have a better suitability for audio and speech applications (Uther 2002; Uther et al. 2007; Uther et al. 2005), and the possibility that mobile devices may offer a more portable and personalised learning solution (Kukulka-Hulme and Shield 2008), it is vital that these new technologies are considered and evaluated appropriately for their effectiveness in relation to language learning applications and the requirements that those kinds of applications might have.

One could argue there is a lack of systematic and comprehensive work to evaluate the suitability of the design of multimedia language learning applications for mobile devices, based on quantitative metrics. Yet, such an endeavour would be tremendously important not only for this project, but also for any learning application using ICT-based multimedia elements (e.g. musical training). Clearly, there are huge implications for the development of commercial products, and the development of comprehensive design guidelines would be a novel and timely contribution to the field. The first step in evaluation is of course to systematically explore the user experience. Here, the concepts of usability and affordances are helpful starting points.

1.1. The concept of 'usability' and relevance of 'affordances'

The evaluation of software necessarily draws us to the broader field of 'usability' (see Nielsen 1993; Shackel 1986, 1991; Noyes 2010 for various definitions that encompass practical and social aspects). The concept of usability has more recently emphasised emotional aspects (e.g. attractiveness and emotions elicited by the applications) and has developed now into a broader framework of 'user experience' (Jordan 1998; Picard 1999). Notwithstanding these affective considerations, the current 'gold standard' defining criteria for usability still remains the definition adopted by the International Organization for Standardization (ISO), which defines three key criteria: effectiveness, efficiency and satisfaction (Faulkner 2000).

Of course, although these aspects are all important to the use of a technology, McGrenere and Ho (2000) suggest that one needs to separate the concept of *usefulness* from usability. Indeed, for learning technologies, this is crucial. There is no point in an application or technology fulfilling usability criteria if the application is not useful as a learning application. For a particular piece of software to be useful, McGrenere and Ho (2000) argue that the technology needs to have an 'affordance' for a particular function. The concept of affordance, put simply, means that the object intuitively and easily lends itself to being used in a specified way by a user (McGrenere and Ho 2000). The development of the concept of affordance and its utility in assisting the design of technology are expanded further below.

The concept of an 'affordance' is anchored within Gibson's view of perception (Gibson 1979), which uses an ecological perspective of perception. Within Gibson's view, the study of the perception of motion needs to be understood within the context of the observer's movement (Greeno 1994). As such, Gibson's theory of perception could be termed a 'situative' theory (Greeno and Moore 1993). Gibson uses the term 'affordance' to refer to the aspects of the environment that contributes to a particular kind of interaction occurring (Greeno 1994). Within Gibson's view, affordances are always seen as relational. For example, the idea that a particular aspect of the environment may have an 'affordance' for a certain ability or action is only understood within the context of an agent that is interacting with that aspect of the environment (Greeno 1994). For example, stairs would afford climbing for an able-bodied person, but not for someone who is confined to a wheelchair.

A further development (and popularisation) of the term 'affordances' came with the work of Norman (1988). In contrast to Gibson, Norman suggests that there are both perceived and actual properties of an object that can lend itself to affordances (e.g. a chair can afford the act of sitting, but the chair can also be stood on; McGrenere and Ho 2000). Another contrast to Gibson is that Norman suggests that affordances are dependent upon the experience, knowledge and culture of the agent, whereas Gibson's assumption is that affordance is independent of the agent's experience, knowledge and culture (McGrenere and Ho 2000). Within Norman's view, affordances can therefore be learnt, whilst within Gibsonian tradition, affordances cannot be learnt.

In more recent years, a distinction between 'perceived' and 'real' affordance has resulted in a useful further development in what could be termed a 'multiple abstractions' view of affordance. Turner (2005) describes

'simple affordances' as being related to the Gibsonian concept and described another layer of abstraction (under which Norman's view of affordance could be grouped) as 'complex affordances'. According to Turner, these two layers of abstraction are distinguished by their context dependence and learnability, with complex affordances being highly context dependent and subject to learning effects. For the current study, a framework developed by Hartson (2003), which specifies further types of affordances, is particularly useful. Hartson identifies four layers of abstraction for affordances: physical, sensory, cognitive and functional, summarised in Table 1.

Within Hartson's framework, the two kinds of affordance that most relate to the Gibsonian and Norman approaches to affordance are the physical and cognitive affordances. Physical affordances, Hartson argues, are the design features of a device or object that lend themselves to a particular action. This might be, for example, having a button that is large enough for users to click on it. Cognitive affordances, on the other hand, are design features that assist the user in identifying how a particular object could be used. For example, a button label may help users to know what happens if they click on it, as in Figure 1. Here, the label and picture signal to the user that they can click on that button to add the item in question to their purchase list.

Hartson (2003) further specifies two additional kinds of affordances: sensory and functional (see Table 1). A sensory affordance, Hartson argues, is having a font size that is large enough to read the label clearly. One could also specify further that sensory affordances could be a product of the software (as in the font size of the actual label). Functional affordances, on the other hand, are design features that help users accomplish a specific task (e.g. a sort function).

It should be noted that the concept of affordances is not uncontroversial (see Albrechtsen et al. 2001; Boyle and Cook 2004; Conole and Dyke 2004; Oliver 2005). However, most of the controversy appears to be related to the difficulty in controlling for the fact that users' affordances are often context-dependent and learnt (be that actively learnt, or implicitly via sociocultural mechanisms). However, one could also argue that the need to consider context and learning effects may simply require the use of experimental designs that allow the study of context in as systematic a way as possible.



Figure 1. Example label that could offer a *cognitive* affordance to complete an online purchase.

Table 1. Adapted summary of different kinds of affordances as specified by Hartson (2003).

Affordance type	Description	Example	Comparison of different terminologies used by Gibson and Norman
Cognitive	Design feature that helps users in knowing something (in relation to the object)	A button label that helps users know what will happen if they click it	Gibson: Perceptual information about an affordance Norman: Perceived affordance
Physical	Design feature that helps users in doing a physical action	A button that is large enough so that users can click on it accurately	Gibson: Affordance Norman: Real affordance
Sensory	Design feature that helps users sense something (especially in relation to cognitive affordances and physical affordances)	A label font size large enough to read easily	Gibson: Not specified Norman: Not specified
Functional	Design feature that helps users accomplish work (i.e. the usefulness of a system function)	The internal system ability to sort a series of numbers (invoked by users clicking on the sort button)	Gibson: Not specified Norman: Not specified

1.2. Applying affordances to design and evaluation of multimedia language learning technologies

Despite the attractiveness of affordances to the field of usability, there has been little systematic use of the affordance concept for design and assessment, except for a couple of notable cases. Before considering the applicability of affordances in multimedia language learning technology, it is useful to consider how affordances might have been already used to inform the design and evaluation of some technological applications. Some important contributions for this project are the studies of Morgan, Butler, and Power (2007) who evaluated affordances of the iPod, DS and Wii for educational purposes and also the work of Churchill, Fox, and King (2012) looking at affordances for the iPad for teachers. Morgan, Butler, and Power's (2007) study of the iPod (which has an identical user interface to the iPhone) has primarily focused on quantitative analyses of physical features. By contrast, Churchill, Fox, and King's (2012) study of the iPad was a qualitative analysis of teachers' use of the iPads. However, within the kind of taxonomy for describing affordance developed by Hartson, these two studies take very different and almost extreme approaches. Morgan's study almost exclusively focuses on the physical affordance side, whereas Churchill

focuses more on functional affordances. Very little consideration is given to sensory or cognitive affordances.

Hartson's (2003) characterisation of sensory and cognitive affordances could be considered a very useful framework for evaluating the affordances of different devices to support applications using multimedia elements because it allows the specification of different aspects of the technology that are relevant to the task. For language learning applications, there are essentially four kinds of tasks that require mastery (listening, speaking, reading and writing), each with their own requirements (Saville-Troike 2006), and it could be argued that for spoken language technologies, it is more important to get the affordances 'right' as the material will not be useful if the rendering of the audio quality is poor. An attempt at mapping on the basic requirements for the different kinds of language learning tasks on mobile technologies according to Hartson's framework has been made in Table 2.

A notable omission from this framework is a consideration of functional affordances. This is because it is difficult to predict functional affordances a priori as this directly ties to the specific type of application and cannot be expressed in generic terms. For example, there can be different kinds of listening programmes that are very useful for aural language, such as audio book or phonetic training programmes. However, one kind of programme

(e.g. audio book) requires a concentrated listening to longer passages, whereas phonetic training is focused on the phoneme level, so programmes concentrate on delivering multiple short trials. Each programme has an entirely different flavour and so therefore consideration of any functional affordances must necessarily be system specific. Therefore, for this study, functional affordances (in terms of actual learning outcomes) are not discussed in detail, although perhaps could be the focus of further research, on a case-by-case basis.

Arguably, the most important types of affordance to psychologists are cognitive and sensory affordances because they are considerations that affect the perception and engagement with the material. In this sense, affordances appear as distinct from, but a necessary precursor to good learning outcomes. Cognitive and sensory affordances are, therefore, of most interest in design issues, as they are aspects that are usually more determined by the software and are also often highly subjective. The physical affordances, on the other hand, generally relate to the physical characteristics of the device. These are generally fixed aspects within a system, usually (but not always) determined by the hardware, but certainly of interest to consider if a 'user-centred' design is considered.

Of course, affordances are generally speaking very difficult to test within controlled experiments because hardware design often determines a priori whether the affordance is there or not (for example, a very small screen will simply not afford to the task of reading as well as a large screen, despite the best attempts of a software designer). As engineers embark on design of hardware, consideration is generally restricted to the physical affordance level, with consideration of cognitive or sensory affordance left to the task of software designers. The difficulty in investigating the role of cognitive or sensory affordances for applications such as language learning technology is that there is a myriad of different kinds of devices that may be used for language learning (iPods, iPads, smartphones, PCs, tablets, etc.). Each have their own physical constraints and it is difficult to match these in a controlled way to inform good practice in user design. However, with the development of the iPad and iPhone, there is a rare opportunity to study affordances for the use of multimedia in language learning. With both types of devices, there are picture quality differences, but audio capabilities and physical interface (button and touch screen) are identical in both cases. A quick comparison of the features of both devices is given in Table 3.

Table 2. Basic requirements for language learning tasks on mobile language technologies in terms of Hartson's (2003) framework of affordances.

Language learning task	Cognitive affordance	Physical affordance	Sensory affordance
Listening	Did the target words or sentences sound comprehensible and a reasonable pace?	Determined by the physical limitations of the device (e.g. frequency response; dynamic range)	Does sound generally sound 'good' on this device? Does it require headphones or speakers to do so?
Speaking	Does the recording function have a clear button (e.g. a microphone icon)? Is there a clear signal to the user where the microphone is actually located?	Determined by the physical limitations of the device (e.g. sampling rate)	Not applicable (unless using a 'foldback' monitor for professional recordings)
Reading	Does the user interface provide clear signals to the user for page turning options, etc.?	Determined by the physical limitations (e.g. screen size and resolution)	Determined by aspects such as font size, backlight and contrast
Writing	Does user interface make it clear when and where the user needs to type text?	Determined by keyboard or text entry input system	Determined by aspects such as cursor position, and speed of entry into the system

1.3. Rationale for current study

Despite the fact that the iPad and iPhone are very similar and applications are often used on both platforms, their

Table 3. Comparison of physical constraints on two Apple devices (values compared from the manufacturer's specifications).

Attribute	iPad2	iPhone4
Size (including screen size)	9.7 in. (diagonal)	3.5 in. (diagonal)
Weight	601 g	137 g
Video quality	H.264 video up to 1080p; 30 frames/s	H.264 video up to 720p; 30 frames/s
Audio quality ^a	20–20,000 Hz	20–20,000 Hz
Display density	132 pixels/in.	360 pixels/in.

^aOther aspects of audio quality (e.g. intensity) were controlled for by using headphones rather than speakers and also using a sound level meter to measure the dB level of output.

cognitive and sensory affordances have not yet been systematically explored. Studying their cognitive and sensory affordances can provide useful information that would inform the future design of spoken language learning technologies. As it has been hypothesised (Uther 2002; Uther et al. 2005) that mobile phones have an affordance for listening to spoken language, it would be useful to compare spoken language learning applications in a mobile phone that has a virtually identical counterpart in a larger device. From the 'cognitive affordance' point of view, one could argue that the very nature of a phone signals functions for speaking and listening to. Such a comparison would also provide a useful framework for the evaluation of these technologies that can assist practitioners conducting usability studies for industry. We also have two software applications with two different kinds of multimedia (audio vs. video) that can be compared: an audio-intensive language learning programme (UCL Vowel Trainer) and a video-intensive language learning programme, focused on grammar ('Learn English for Taxi Drivers' by British Academy). Both software applications would be tested on both types of devices (iPhone and iPad).

Based on the literature reviewed, four hypotheses emerge:

- (1) The first hypothesis (H1) is that iPhones have a better sensory affordance for listening to sound (due to their positioning as a phone), whereas iPads might have a better sensory affordance for reading and watching video (due to the screen size). Sensory affordance is tested by users' ratings of sound and video quality, independent of the software application.
- (2) The second hypothesis (H2) is that in terms of cognitive affordance (i.e. perceptions of qualities within the context of specific software), there are two possible predictions. On the one hand, it could be predicted that iPads afford better for the video-based application as it has a larger screen and iPhones

afford for the audio language learning application as it is seen as better for listening (again, because a phone is commonly used for speaking and listening). One could also plausibly predict that both types of applications might afford equally well because the iPad is seen by users as a better 'educational' tool (due to its positioning as a more general multimedia device).

- (3) The third hypothesis (H3) is that high ratings for listening and watching would also be coupled with the evidence of higher overall usability as indexed by standardised usability measures.
- (4) Finally, the fourth hypothesis (H4) is that there are differences in how native language English content (e.g. audio books) might be rated for quality by non-native speakers compared to native speakers of English, with the expectation that native English speakers rating English content rate their native content higher.

2. Methods

2.1. Participants

Forty-one participants were recruited from the University of Surrey campus and ethical clearance was obtained. For the native speaking group ($n = 21$), the participants were from an English monolingual environment. For the non-native speaking group ($n = 20$), these participants were recruited from the School of Languages and were students who had recently migrated to the UK or were on an exchange visit to the university. All participants had a sufficiently good command of written and spoken English (as indexed by a linguistic background questionnaire that measured proficiency on a scale of beginners, intermediate and advanced/native users of English) in order to understand the task and the informed consent material. Nineteen participants were iPhone users, 16 participants were Android Smartphone users and 4 participants used a more traditional mobile phone with very limited multimedia capability or Internet access.

2.2. Materials

The study tested participants on tasks using an iPhone version 4 and an iPad version 2. Both devices were also loaded with the UCL Vowel Trainer and the British Academy Learn English for Taxi Drivers' software (see Figure 2 for screenshots of each).

A pair of Sennheiser headphones was used to standardise the sound output and sound level was kept constant by taking measurements using a sound level meter.

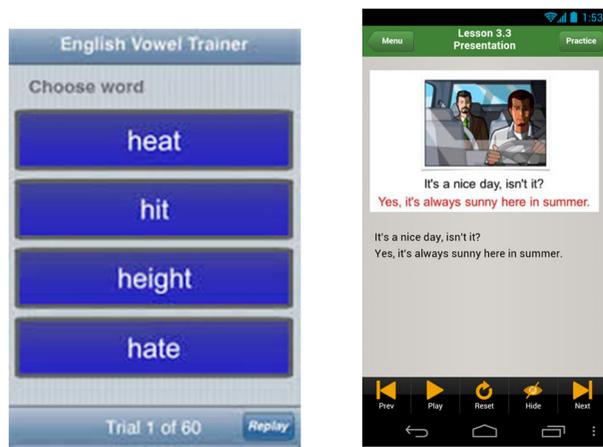


Figure 2. Screenshots of interface from both software applications. The left hand panel is the UCL Vowel Trainer and the right hand panel is the Learn English for Taxi Drivers software. Screenshots are taken from the manufacturer's promotional material and do not reflect actual resolution on the same screen.

Several questionnaires were used. The first was a general demographic and language background questionnaire gathering data about age, gender, mobile device ownership and native language background. There were bespoke questionnaires 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 the worst). A task-adapted Questionnaire for User Interface Satisfaction (QUIS) was used (Chin, Diehl, and Norman 1988) to measure users' overall satisfaction. The QUIS also includes sub-scales on screen interface, terminology, learning and system capabilities.

2.3. Design

The study was run as a 'mixed' design, with native language (native vs. non-native English) as group factor ($n = 21$ for native English and $n = 20$ for non-native group) and device type (iPad vs. iPhone) and software type (multimedia rich including audio, video and graphics vs. audio rich only) 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, language background and mobile phone usage were also collected prior to the study commencing using another questionnaire.

Participants then completed a set of tasks on the iPhone and iPad and the order of presentation of each task on each device was counterbalanced. As an index of sensory affordances, participants first rated sound

and video quality of generic content on the iPad and iPhone. For sound quality, two audio samples were played: (1) a speech sample: a standardised, short, 15-s passage from an audio book (No. 1 Lady's Detective Agency). (2) A musical sample: a short, 15-s sample of music from Yo-Yo Ma's rendition of Bach's Cello Suite #1 in 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-s sample of a high-definition video (National Geographic's documentary on the 'Mutant blond penguin in 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 language learning applications (a combination of a short (minute long) free play and a prescribed sample task of the application) and were then asked to rate the software and device. The overall user experience was rated using the Questionnaire for User Interface Interaction (Chin et al. 1988), which covers key areas including overall reaction to the software; screen characteristics; terminology and system information; and learning and system capabilities. Participants were also asked to explicitly rate the audio and video quality as well as affordance of each device being used for each software application.

3. Results

The data were analysed using a mixed multivariate ANOVA, as well as correlations of affordance ratings with QUIS. There were three main datasets: sensory affordance, cognitive affordance and user satisfaction, which were analysed in turn. *Post hoc* tests were also carried out on ANOVAs with iPhone usage as a grouping variable in order to control for device familiarity. This did not yield any change to the results presented here.

3.1. Sensory affordance data

The subjective comparison of audio and video quality of generic samples across iPad and iPhone was taken as an index of sensory affordances for each device. The results are detailed below.

For audio quality, three types of sound were rated for user-perceived quality (audio book, music sample or audio part of the video clip). Results showed there was a main effect of device on audio rating, with the iPad having higher audio quality ratings regardless of the sound type ($F_{(1,39)} = 11.885$, partial $\eta^2 = .234$, $p < .001$; see Figure 3).

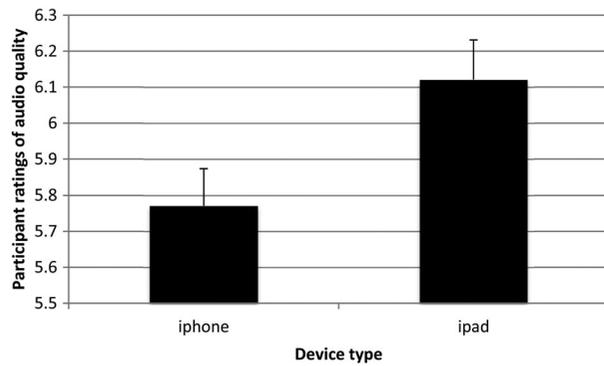


Figure 3. Participant-rated audio quality of generic sound samples across both devices. Error bars show standard errors.

Results also showed a main effect of sound type ($F_{(1,39)} = 9.842$, partial $\eta^2 = .201$, $p < .001$; see Figure 4). Pairwise comparisons revealed, however, that it was the audio aspect of the video clip that was rated lower in general compared to the other audio (music or audio book) clips ($p < .05$). There were no significant differences between user's rating of audio quality between the audio book and music sample (see Figure 4).

In terms of picture quality, results showed there was a main effect of device on picture rating, with the iPad in general having higher picture ratings ($F_{(1,39)} = 20.769$, partial $\eta^2 = .347$, $p < .001$; see Figure 5).

Overall, contrary to H1, the iPad showed better sensory affordance for both sound and video. Interestingly, no language group main effects or interactions with language group were found. This suggests that H4 predictions of differences in native vs. non-native speaker ratings were not supported.

3.2. Cognitive affordance data

The subjective comparison of audio and video quality of each language learning software across iPad and iPhone

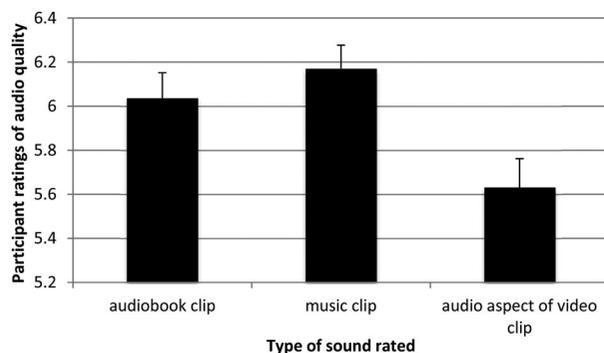


Figure 4. Participant-rated audio quality over three kinds of audio samples. Error bars show standard errors.

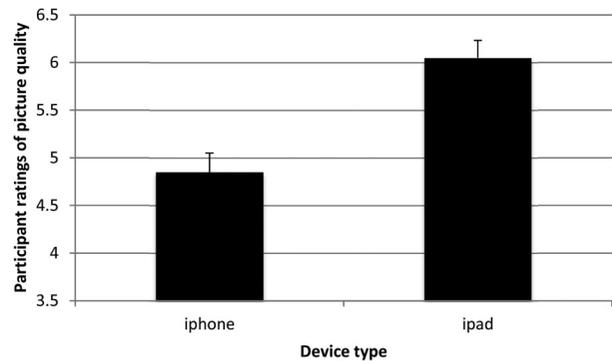


Figure 5. Participant-rated picture quality across both iPhone and iPad devices. Error bars show standard errors.

was taken as an index of cognitive affordances for each device. The results are detailed below.

Results showed there was a main effect of device only on picture rating for the video clip, with the iPad having higher picture ratings across both software apps ($F_{(1,39)} = 4.922$, partial $\eta^2 = .112$, $p < .05$; see Figure 6). The audio quality ratings did not differ between the two devices, suggesting that both types of device afforded equally well for the listening experience.

There was a main effect of software on the rating for audio quality, with the 'Learn English for Taxi Drivers' (i.e. text-based) app having higher audio ratings across both devices compared to the 'Vowel Trainer' (i.e. audio-based) app ($F_{(1,39)} = 9.397$, partial $\eta^2 = .2$, $p < .05$; see Figure 7).

There was also a significant device by software interaction for picture quality, with the 'Learn English for Taxi Drivers' (i.e. text-based) app having quite different picture ratings across the two devices compared to the 'Vowel Trainer' (i.e. audio-based) app ($F_{(1,39)} = 9.794$, partial $\eta^2 = .2$, $p < .05$; see Figure 8). One can see that participant ratings were similar for the Vowel Trainer across both devices, whereas the 'Learn English for

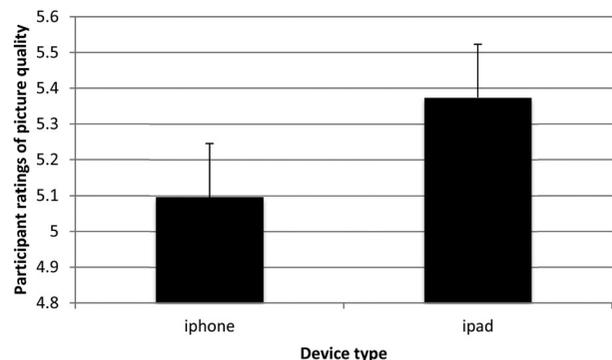


Figure 6. Participant-rated picture quality for both devices across both types of language learning software. Error bars show standard error.

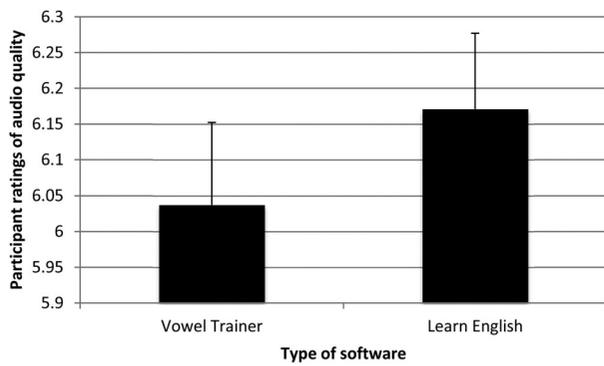


Figure 7. Mean participant-rated audio quality for the two language learning software applications. Error bars show standard error.

Taxi Drivers' software app had a much lower rating for the iPhone compared to its iPad rating.

Overall, the second hypothesis (H2), which predicted that larger screen size would afford richer media content, was supported. As for the sensory affordance data, no language group main effects or interactions with language group were found. This suggests again that H4 predictions of differences in native vs. non-native speaker ratings were not supported.

3.3. Subjective ratings of suitability of application to device

Ratings of participants' perceptions of suitability of each application to each device were also measured, along with their perceived likelihood of future use if they owned that application. Here, there were no significant main effects of group, nor device, nor software. However, there was a significant device x software interaction for both the 'suitability of device' and 'likelihood of future use' ratings.

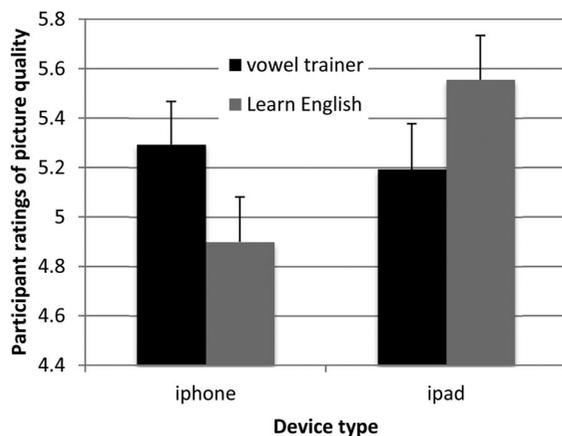


Figure 8. Participant-rated picture quality for both language learning applications. Error bars show standard error.

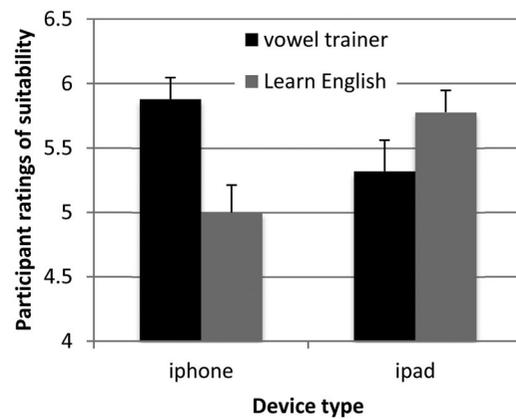


Figure 9. Participant-rated suitability of each software for each device. Error bars show standard error.

For the rated suitability for device, the interaction showed that for the iPad, each software tended to be rated more similarly, whereas for the iPhone, the 'Learn English for Taxi Drivers' (visually based) software was rated more poorly in comparison to the Vowel Trainer software ($F_{1,39} = 14.296$, $p < .001$, partial $\eta^2 = .268$; see Figure 9).

For the rated likelihood for future use, the interaction was similar to the previous effect on rated suitability, which showed that for the iPad, each software tended to be rated similarly, whereas for the iPhone, the Learn English for Taxi Drivers (visually based) software was rated more poorly in comparison ($F_{1,39} = 7.896$, $p < .05$, partial $\eta^2 = .168$; see Figure 10).

3.3.1. Analysis of QUIS ratings on each software

The QUIS ratings were broken down into overall scores and sub-scales relating to screen-related elements (screen), ease of learning, terminology and system capability. All measures were significantly different between the two software applications ($F_{1,39} = 12.019$, $p < .01$,

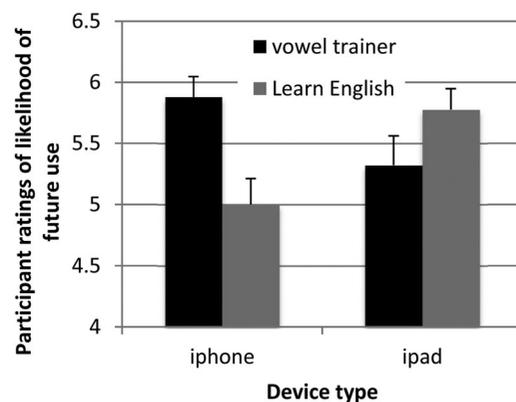


Figure 10. Participant-rated likelihood of future use on each platform for each application. Error bars show standard error.

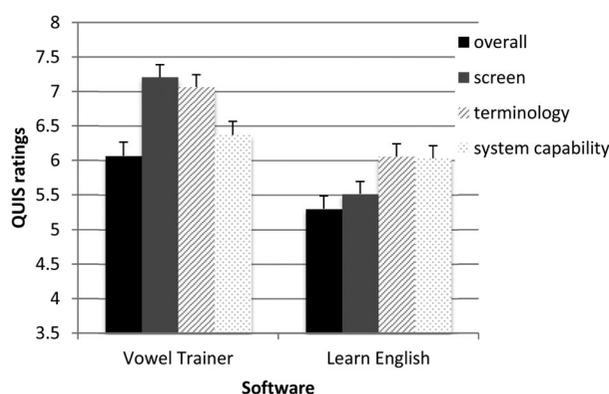


Figure 11. QUIS user interface ratings for each software.

partial $\eta^2 = .236$ for overall scale; $F_{1,39} = 76.66$, $p < .01$, partial $\eta^2 = .663$ for screen elements; $F_{1,39} = 36.335$, $p < .01$, partial $\eta^2 = .482$ for terminology; $F_{1,39} = 6.646$, $p < .05$, partial $\eta^2 = .142$ for system capability), except for the 'ease of learning' sub-scale which did not statistically differ. In general, the Vowel Trainer (speech-based) application scored more highly, see Figure 11.

There was also a significant software by group interaction on the terminology sub-scale ($F_{1,39} = 4.172$, $p < .05$, partial $\eta^2 = .097$), whereby the ratings for the two software applications were more similar for the non-native speakers compared to quite different ratings for the two applications in the native speaking group (see Figure 12). However, the effect size here was quite small. Although this finding was not directly linked to the main hypotheses, it was the only instance where there were any statistically significant differences between the two groups. It suggests that the native speakers may have picked up some wording flaws in the user interface for the software which had more grammatically complex material, which shows that they were indeed reading native content slightly differently from non-native speakers.

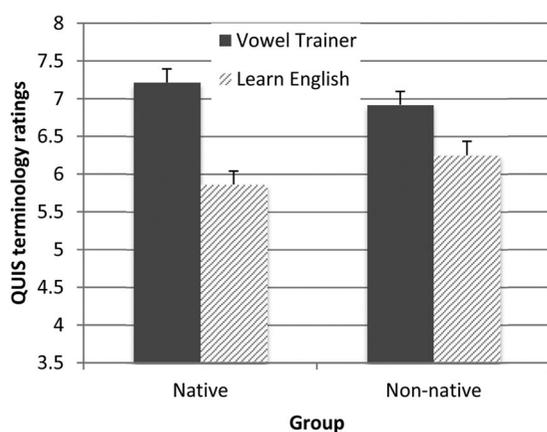


Figure 12. QUIS terminology ratings for native and non-native speakers in each software application.

Although one could predict that affordance data in relation to the software might be affected by overall differences in the ease of use of the software, the evidence does not support this. If anything, a superior rating is given to the application that tends to fare worse on audio and picture quality ratings (i.e. Vowel Trainer). Correlations between rated audio and picture quality and QUIS results also did not show a clear relationship between quality and overall user experience ratings. Although for the Learn English for Taxi Drivers (picture/reading-based software) software the QUIS ratings correlated well with rated audio and picture quality (Pearson's $r = 0.53$, $p < .05$ and $r = .57$, $p < .05$, respectively), the QUIS ratings did not correlate uniformly well with the Vowel Trainer (speech-based software). Here, the Vowel Trainer audio quality ratings only correlated with QUIS ratings (Pearson's $r = .53$, $p < .05$) but did not correlate significantly with picture quality ratings (Pearson's $r = .35$, $p > .05$). Overall, the third hypothesis (H3) was not supported – suggesting that subjective affordances do not correlate with the final user ratings.

4. Discussion

The ratings of picture quality (an index of sensory affordance for reading/watching) were in a direction that was entirely predictable. With larger screen size, the video content was rated better than that on the smaller screens. By contrast, the rated audio quality showed a slightly different picture. For the 'sensory affordance data' (data from generic sound and picture clips, not within a language learning software), the effects on audio were in a direction that was different from that originally predicted. Originally, it was suggested that audio content would be perceived as better on the smartphone device (in the sense that people would be presumably used to listening to audio content as a phone or music player). However, this prediction was not supported by the data, which, if anything, rated audio quality played on the iPad higher than that of the iPhone. Interestingly, the effect was not neutral (i.e. no difference in perception of audio quality), despite both devices rendering the same physical output. People were obviously rating the audio quality differently. It is of interest to examine why the effect might be in an opposite direction to that originally predicted.

In explaining the sensory affordance findings for audio, it may be because affordances for iPhone might have been overridden by other (contextual) effects. For example, it could be that the 'Smartphone' is not as 'phone-like' as it used to be. Today's devices are often not so much seen as a dichotomous category of being either a phone or a multimedia device, but rather on a continuum. Within

that continuum, it might be that the smartphone is seen as closer to a tablet device than a traditional mobile phone. On the other hand, tablet devices within the continuum of large vs. small multimedia devices might probably be seen as more computationally powerful and, therefore, a 'superior' device. In this way, a larger device might be perceived to offer better sound quality despite the fact that the physical output is the same.

Equally, it may have also been simply due to difficulties in characterising the nature of 'sensory affordances' for listening. It is not entirely clear how a 'sensory affordance for listening' would be specifically cued. Tracing the literature and theories on affordances, theories are rooted within visual perception and Gibsonian tradition (e.g. the affordance of a staircase for climbing). Yet, there is little work on what affordances for *listening* might entail.¹ The appearance of an object as a phone would signal to the user (as a result of experience) that this is a device that would be spoken into and listened from (whether for conversation or for music). However, it is not clear that it would necessarily 'afford' a listening experience that is superior to a larger multimedia device. In the same way that two sets of stairs may equally afford to the user that they could be climbed up, the experience of each may be the result of other factors (e.g. one might be more narrow and steep than the other staircase). Clearly, the appearance of a phone serves as a visual cue for auditory interaction, but that does not mean that it would be a 'superior' device to other devices that would also afford for listening.

Within the context of the actual software, one sees a slightly different picture with respect to the indices of subjective audio quality. Here, 'cognitive affordance' is indexed by how the sensory input is used/interpreted within the particular software and suggests that sound quality does not appear to be directly determined by device size. This discrepancy between sensory and cognitive affordance data for audio ratings may be explained by differences in content types for cognitive vs. sensory affordance data. For sensory affordance (generic media) ratings, two of the three samples tested for audio quality were taken from audio-only clips (with the third taken from a video clip which contained both audio and picture). On the other hand, for cognitive affordance ratings, all audio ratings were taken from multimedia clips (i.e. contained both visual content and sound). This explanation is supported by data from the audio rating of generic content showing that the audio-only clips (music and audio book) were in general rated better than the audio portion of the video clip. These results suggest that the presence of picture content influences the rating of audio content. Future studies could explore this further and in particular whether focusing the user on

'intelligibility' or 'clarity' of the speech (as opposed to generic sound quality) may yield a slightly different result.

There were no noteworthy differences in native/non-native ratings of sound quality of speech recordings (be they software related or audio books) as initially predicted. This could be due to the fact that the non-native speakers were not especially low in proficiency (most were at the 'intermediate' level rather than 'basic'), or it could be that native speakers' perception of 'sound quality' is independent of 'goodness' (which would have undoubtedly differed, being dependent upon categorical perception). Nonetheless, there was a small, incidental finding that 'terminology' was rated differently in the text-based software by the two groups from that in the more speech-based software. This would suggest that the native speakers were reading native content differently from non-native speakers.

In terms of how affordance ratings affect user perceptions, it appears that picture quality, in particular, seemed to impact on rated suitability of software to the device and also the likelihood of future use as directly rated by the participants in the bespoke rating scales. There were hints in the qualitative reports from the users that this apparent discrepancy was due to the fact that they were willing to trade off slight declines in quality to gain more convenience and portability. Comments such as 'simple interface' and 'only useful on a small screen' were made. Hence, it would appear that overall impressions of standard 'usability' metrics do not always translate to decisions to use the software less, particularly for mobile devices. Here, other considerations such as convenience and mobility 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 et al. 2007; Orr 2010).

Gibson's original theory suggested that affordances are perceived directly and do not require mediation or internal processing by the perceiver. However, this (ecological) view of perception does contrast with the cognitive views of perception within current psychological approaches, which suggests that environmental properties of an object (height, width, etc.) are not evaluated on extrinsic, absolute scales with standard units of measurements but rather are perceived on intrinsic scales relative to the perceiver-actor's own physical characteristics (e.g. own height and own width; see Regia-Corte, Marchal, and Lecuyer 2010 for a review). This idea that affordances are not intrinsically determined by physical properties would explain the context-dependent effects seen in this study.

In this way, it is also entirely possible that participants' experience of mobile devices affected the results.

Although data were gathered about which phone they used (and did not yield any significant differences in the data), more fine-grained data might have been useful to distinguish between heavy users and light users. In particular, it would have been useful to know how often they used their phones to listen to audio content and also how often they watched video content. As stated earlier, it is also unclear whether users viewed the iPhone within the category of 'phone' or instead within a more generic category of 'multimedia mobile devices' which would cover tablets as well as smartphones, with small tablets such as the iPad mini being an 'intermediary' device. Hence, a 'gradation' of sound quality according to size may be seen, although this may be dependent on user experience with such devices.

These data nonetheless have clear practical applications for industry. The findings show that subjective perception of audio is affected by context and is independent of the physical constraints (at least with headphone use, as was done with this study). It would appear that the perception of sound quality might be affected by perceived 'powerfulness' of a device, but this may be the case mainly for audio-only content, and certainly more research is needed to confirm this. Secondly, it would appear that the importance of perceived inferior audio quality might be downplayed in the context of other factors such as the presence of video content or considerations about its relative importance in light of other trade-offs (convenience and portability), for example. These data may also be useful not only in the design of language learning applications, but also in other learning applications where audio-based content is important (e.g. music learning applications).

Note

1. It should be emphasised that an affordance for auditory perception (listening) is distinct from 'sound affordances', where auditory cues signal action (e.g. engine sounds indicating it is time to shift gears; see Stanton and Edworthy 1998 for a review).

Acknowledgements

The authors thank Dr Paul Iverson of University College London for the use of the Vowel Trainer in this study and Dr James Uther for technical assistance throughout this project. Preliminary data from this study were published in the proceedings of IADIS Mobile Learning Conference, March 2015.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Maria Uther  <http://orcid.org/0000-0002-7948-302X>

References

- Albrechtsen, H., H. H. K. Andersen, S. Bødker, and A. M. Pejtersen. 2001. *Affordances in Activity Theory and Cognitive Systems Engineering Learning*. Roskilde: Riso National Laboratory, Centre for Human-Machine Interaction (Riso-R-1287). <http://www.riso.dk/rispubl/SYS/ris-r-1287.htm>.
- Boyle, T., and J. Cook. 2004. "Discussion Understanding and Using Technological Affordances: A Commentary on Conole and Dyke." *ALTJ* 12 (3): 295–299.
- Bradlow, Ann R., R. Akahane-Yamada, D. B. Pisoni, and Y. Tohkura. 1999. "Training Japanese Listeners to Identify English /r/ and /l/: Long-term Retention of Learning in Perception and Production." *Perception and Psychophysics* 61 (5): 977–985.
- Bradlow, A. R., D. B. Pisoni, R. Akahane-Yamada, and Y. Tohkura. 1997. "Training Japanese Listeners to Identify English /r/ and /l/: IV. Some Effects of Perceptual Learning on Speech Production." *Journal of the Acoustical Society of America* 101 (4): 2299–2310.
- Chin, J. P., V. A. Diehl, and K. L. Norman. 1988. "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface." In E. Soloway, D. Frye, and S. B. Sheppard (Eds.), *ACM CHI'88 Proceedings*, Vol. 218, 213–218. New York: Association for Computing Machinery.
- Chinnery, G. M. 2006. "Emerging Technologies: Going to the MALL: Mobile Assisted Language Learning." *Language Learning & Technology* 10 (1): 9–16.
- Churchill, D., B. Fox, and M. King. 2012. "Study of Affordances of iPads and Teachers' Private Theories." *International Journal of Information and Education Technology* 2 (3): 251–254.
- Conole, G., and M. Dyke. 2004. "What Are the Affordances of Information and Communication Technologies?" *ALTJ* 12 (2): 113–124.
- Cook, V. 2008. *Second Language Learning and Language Teaching*. 4th ed. London: Hodder Education.
- Faulkner, X. 2000. *Usability Engineering*. Basingstoke: Palgrave Macmillan.
- Fisher, T., R. Pemberton, M. Sharples, H. Ogata, N. Uosaki, P. Edmonds, A. Hull, et al. 2009. "Mobile Learning of Vocabulary from Reading Novels: A Comparison of Three Modes." In *mLearn 2009*, Orlando, FL.
- Giannakopoulou, A., M. Uther, and S. Ylinen. 2013. "Enhanced Plasticity in Spoken Language Acquisition for Child Learners: Evidence from Phonetic Training Studies in Child and Adult Learners of English." *Child Language and Teaching Therapy* 29 (2): 201–218.
- Gibson, J. 1979. *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
- Greeno, J. G. 1994. "Gibson's Affordances." *Psychological Review* 101 (2): 336–342. New York: Association for Computing Machinery.
- Greeno, J. G., and J. L. Moore. 1993. "Situativity and Symbols: Response to Vera and Simon." *Cognitive Science* 17 (1): 49–59.

- Guerrero, L. A., S. Ochoa, and C. Collazos. 2010. "A Mobile Learning Tool for Improving Grammar Skills." *Procedia Social and Behavioral Sciences* 2 (2): 1735–1739.
- Hartson, R. 2003. "Cognitive, Physical, Sensory, and Functional Affordances in Interaction Design." *Behaviour Information Technology* 22 (5): 315–338.
- Hubbard, P. 1987. "Language Teaching Approaches. The Evaluation of CALL Software and Design Implications." In W. Smith (Ed.), *Vol. 1 of Modern Media in Foreign Language Education: Theory and Implementation*, 227–254. Lincolnwood, IL: National Textbook Co.
- Iverson, Paul, and B. G. Evans. 2009. "Learning English Vowels with Different First-language Vowel Systems II: Auditory Training for Native Spanish and German Speakers." *Journal of the Acoustical Society of America* 126 (2): 866–877.
- Iverson, P., V. Hazan, and K. Bannister. 2005. "Phonetic Training with Acoustic Cue Manipulations: A Comparison of Methods for Teaching English /r / -/l/ to Japanese Adults." *Journal of the Acoustical Society of America* 118 (5): 3267–3278.
- Jordan, P. W. 1998. *An Introduction to Usability*. London: Taylor & Francis.
- Kukulka-Hulme, A. 2005. "The Mobile Language Learner – Now and in the Future." In *Fran Vision till Praktik Language Learning Symposium*. Umea: Umea University. <http://www2.humlab.umu.se/symposium2005/program.htm>.
- Kukulka-Hulme, Agnes, and L. Shield. 2008. "An Overview of Mobile Assisted Language Learning: From Content Delivery to Supported Collaboration and Interaction." *ReCALL* 20 (3): 271–289.
- Lai, C.-H., J.-C. Yang, F.-C. Chen, C.-W. Ho, and T.-W. Chan. 2007. "Affordances of Mobile Technologies for Experiential Learning: The Interplay of Technology and Pedagogical Practices." *Journal of Computer-Assisted Learning* 23: 326–337.
- Lively, S. E., J. S. Logan, and D. B. Pisoni. 1993. "Training Japanese Listeners to Identify English /r/ and /l/. II: The Role of Phonetic Environment and Talker Variability in Learning New Perceptual Categories." *Journal of the Acoustical Society of America* 94 (3 Pt 1): 1242–1255.
- Lively, S. E., D. B. Pisoni, R. A. Yamada, Y. Tohkura, and T. Yamada. 1994. "Training Japanese Listeners to Identify English /r/ and /l/. III. Long-term Retention of New Phonetic Categories." *Journal of the Acoustical Society of America* 96 (4): 2076–2087.
- McGrener, J., and W. Ho. 2000. "Affordances: Clarifying and Evolving a Concept." In *Proceedings of Graphics Interface 2000*, 179–186. Montréal: Canadian Information Processing Society.
- Miller, J. D., C. S. Watson, D. J. Kistler, J. E. Preminger, and D. J. Wark. 2008. "Training Listeners to Identify the Sounds of Speech: II. Using SPATS Software." *The Hearing Journal* 61 (10): 29–33.
- Morgan, M., M. Butler, and M. Power. 2007. "Evaluating ICT in Education: A Comparison of the Affordances of the iPod, DS and Wii." In *Proceedings ASCILITE*. Singapore: Centre for Educational Development, Nanyang Technological University, 2–5 December. <http://www.ascilite.org.au/conferences/singapore07/procs/>.
- Nielsen, J. 1993. *Usability Engineering*. London: Academic Press.
- Norman, D. 1988. *The Design of Everyday Things*. New York: Basic Books.
- Noyes, J. 2010. "Design and Use of Work Technology." In N. Chmiel (Ed.), *An Introduction to Work and Organisational Psychology*, 2nd ed., 151–168. Oxford: Blackwell Publishing.
- Oliver, M. 2000. "An Introduction to the Evaluation of Learning Technology." *Educational Technology & Society* 3 (4): 20–30.
- Oliver, M. 2005. "The Problem with Affordance." *ELearning* 2 (4): 402–413.
- Orr, G. 2010. "A Review of Literature in Mobile Learning: Affordances and Constraints." In *Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE), 6th IEEE International Conference*, April, 107–111. Kaohsiung: IEEE Computer Society.
- Picard, R. 1999. "Affective Computing for HCI." In *Proceedings HCI International (8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces*, Vol. 1, 829–833. Hillsdale, NJ: L. Erlbaum Associates, Inc.
- Plass, J. L. 1998. "Design and Evaluation of the User Interface of Foreign Language Multimedia Software: A Cognitive Approach." *Language Learning & Technology* 2 (1): 40–53.
- Regia-Corte, T., M. Marchal, and A. Lecuyer. 2010. "Can You Stand on Virtual Grounds? A Study on Postural Affordances in Virtual Reality." In *IEEE Virtual Reality Conference (VR)*, March, 207–210. Waltham, MA: IEEE Computer Society.
- Saville-Troike, M. 2006. *Introducing Second Language Acquisition*. 1st ed. Cambridge: Cambridge University Press.
- Shackel, B. 1986. "Ergonomics in Design for Usability." In *Proceedings of the Second Conference of the British Computer Society, Human Computer Interaction Specialist Group on People and Computers: Designing for Usability*, edited by M. D. Harrison, and A. F. Monk, 44–64. Cambridge: Cambridge University Press.
- Shackel, B. 1991. "Usability – Context, Framework, Definition, Design and Evaluation." In *Human Factors for Informatics Usability*, edited by Brian Shackel and Simon J. Richardson, 21–37. New York: Cambridge University Press.
- Stanton, N., and J. Edworthy. 1998. "Auditory Affordances in the Intensive Treatment Unit." *Applied Ergonomics* 29 (5): 389–394.
- Strei, G. 1983. "Format for the Evaluation of Courseware Used in Computer-assisted Language Instruction (CALI)." *CALICO* 1 (2), 43–46.
- Thornton, P., and C. Houser. 2005. "Using Mobile Web and Video Phones in English Language Teaching: Projects with Japanese College Students." In *Directions in CALL Experience Experiments Evaluation*, edited by B. Morrison, C. Green, & G. Motteram, 207–224. Hong Kong: Hong Kong English Language Centre, Hong Kong Polytechnic University.
- Turner, P. 2005. "Affordance as Context." *Interacting with Computers* 17 (6): 787–800.
- Uther, M. 2002. "Mobile Internet Usability: What Can "Mobile Learning" Learn from the Past?" In *Proceedings IEEE International Workshop on Wireless and Mobile Technologies in Education*, 174–176. Vaxjo: IEEE Computing Society.
- Uther, M., J. Uther, P. Athanasopoulos, P. Singh, and R. Akahane-Yamada. 2007. "Mobile Adaptive CALL (MAC): A Lightweight Speech-based Intervention for Mobile Language Learners." *Interspeech*. Antwerp: ISCA.

- Uther, M., I. Zipitria, J. Uther, and P. Singh. 2005. "Mobile Adaptive CALL (MAC): A Case-study in Developing a Mobile Learning Application for Speech/audio Language Training." In *IEEE Wireless and Mobile Technologies in Education 2005*. Tokushima: IEEE Computing Society.
- Watson, C. S., J. D. Miller, D. Kewley-Port, L. E. Humes, and F. L. Wightman. 2008. "Training Listeners to Identify the Sounds of Speech: I. A Review of Past Studies." *The Hearing Journal* 61 (9): 26–31.
- Ylinen, S., M. Uther, A. Latvala, S. Vepsäläinen, P. Iverson, R. Akahane-Yamada, and R. Näätänen. 2010. "Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English." *Journal of Cognitive Neuroscience* 22 (6): 1319–1332.