Research and Practice in Technology Enhanced Learning
Vol. 4, No. 1 (2009) 7–32

MOBILE DEVICES FOR LANGUAGE LEARNING: MULTIMEDIA APPROACHES

SAMUEL R. H. JOSEPH
Laboratory for Interactive Learning Technologies
Department of Information & Computer Sciences
University of Hawaii, 1680 East West Road, POST 309
Honolulu, HI 96822, USA
srjoseph@hawaii.edu
http://lilt.ics.hawaii.edu

MARIA UTHER
Centre for Cognition & NeuroImaging
School of Social Sciences
Brunel University, Uxbridge
Middlesex UB8 3PH, United Kingdom
Maria.Uther@brunel.ac.uk
http://people.brunel.ac.uk/∼hsstmmu/

This paper examines the use of mobile devices for language learning. In particular we consider how different multimedia and interface modalities can be used to facilitate mobile language learning. The use of multimedia is considered within the context of Second Language Acquisition (SLA) and Computer Assisted Language Learning (CALL) theories. In particular, we examine the Interaction Framework, Schemata Theory, SLA Universal Grammar related theories and Multimedia Learning Theory. Two case studies serve to illuminate: the audio-based training of a system called MAC and the image-based training of another called PhotoStudy. We also suggest evaluation techniques that should be used to test the application of these theories to Mobile Assisted Language Learning (MALL).

Keywords: Mobile devices; language learning; multimedia.

1. Introduction

The development of mobile and wireless technologies has opened up a huge array of possibilities in the domain of language learning. There have been trials of mobile-assisted language learning (MALL) applications since 2001 (e.g. Stanford Learning Lab, 2001). Chinnery (2006) and Kukulska-Hulme & Shield (2008) both provide reviews of many MALL developments since then, and although there have been some interesting uses of multimedia on PDAs and iPods (e.g. Thornton & Houser, 2003; Garcia Cabrere, 2002; Belanger, 2005; Kukulska-Hulme, 2005), the focus has generally been on delivering basic features such as vocabulary learning and quiz drills in text format over mobile phones.

However, successful language acquisition involves developing far more complex skills than those used in simple vocabulary drills. Broadly, the aims of language acquisition can be classified into either receptive or productive language use, which can be further subdivided into one of 4 main categories (depending on whether written or oral): speaking, listening, reading and writing (Saville-Troike, 2006). Developing good receptive and productive language skills naturally involves elements such as the acquisition of grammar and vocabulary as well as more abstract elements such as subject matter and cultural knowledge (Gathercole & Conway, 1988).

Recent approaches to MALL (e.g. Joseph *et al.*, 2005; Uther *et al.*, 2005a,b,c, 2007; Yang *et al.*, 2005) have started to move beyond simple text drills. This paper reviews the implications of the latest Computer Assisted Language Learning (CALL) and Second Language Acquisition (SLA) theories for MALL and examines some case studies that highlight the role of these theories in the use of mobile multimedia. Other theoretical implications have yet to be followed up in the MALL field. We consider the challenges that must be faced in order to adopt them, and the evaluation techniques that must be employed to validate their effectiveness. We conclude with a discussion of future media possibilities that go beyond the simple audio and visual modalities.

2. Learning Theory & Mobile Language Learning (MALL)

In this section, we will discuss a number of pedagogical theories in turn and look at how they influence or affect development of MALL systems. In doing so we need to think not only in terms of potential instruction of the learner, but also in terms of potential intervention or scaffolding in their daily life. The unique aspect of a MALL system is that its operation is not confined to a classroom, or the part of the day when the learner happens to be at a computer. An important question for learners is "how do I spend my time most efficiently in order to achieve second language ($L2$) competence?", and MALL extends the CALL domain into everyday activities. One way to look at the different theories is to think about their implications for instruction, but instruction as a concept is probably best suited to a classroom environment. Various pedagogical theories have recommendations for instruction, but replicating the classroom experience on a mobile device is probably ill-advised due to device input/output limitations (e.g. small screen and keyboard). In order for theories to inform MALL they need to answer the questions: How should the learner spend their day in order to optimize learning, and how might a mobile device help?

Two theories of learning that strongly influence the teaching of language are behaviorism (for a recent review on its influence on SLA, see VanPatten & Williams, 2007) and constructivism (for a recent overview of its influence on the language classroom, see Schcolnik, Kol & Abarbanel, 2006). The behaviorist tradition sees the learner as a starting from a blank slate and treats the learning process as a mechanistic one that can be influenced by positive and negative reinforcements. By contrast, the constructivist tradition argues that a learner actively constructs new ideas based around their existing knowledge. In addition, constructivism emphasizes the importance of collaboration in learning, or knowledge construction as a social process. Arguably the mobile context has much to offer in the way of supporting learner interaction, collaboration, and the co-construction of knowledge. However, as is generally the case with most computer-based systems, it is all too easy for the designer of a mobile language learning application to opt for a behaviorist-oriented or programmed instruction approach.

2.1.1. *Grammar focused SLA theories*

The field of Second Language Acquisition (SLA) includes a number of theories relating to the acquisition of morphology (structure and form of words) and syntax (grammatical rules), as well as a little phonology, and occasionally that of lexis (lexical store of words and phrases) and pragmatics (social context) (Levy, 1997). The first major empirical studies of SLA performed in the 1970s did not support the use of a behaviorist approach in teaching. Studies showed that positive and negative reinforcement, or error correction, did not always improve learner performance (Van Patten & Williams, 2007). Although behaviorism is largely discredited in the SLA domain, it would be a mistake to characterize all SLA theories as constructivist, particularly those that propose a central role for a Universal Grammar (UG); a linguistic theory that postulates an innate underlying grammar that is shared by all human languages. SLA UG theory (White, 2007) asserts that grammatical knowledge cannot be taught, but speculates that carefully tuned instruction might help set particular grammatical parameters; though attempts at this have led to mixed results (White, 2003). The central evidence for a Universal Grammar is a "poverty of stimulus" argument, which suggests that the linguistic input a learner is exposed to is insufficient for the learner to derive certain grammatical features that are consistently observed in native speakers.

Connectionist researchers such as Elman *et al.* (1996) suggest that the Universal Grammar concept is based on a too-strict, worst case model of grammar, and other recent SLA theories have posited a more implicit role for UG (e.g. Caroll, 2007). However, Processability theory (Pieneman, 2007) posits that learners cannot learn grammatical rules outside of a particular order or hierarchy. Processability theory and SLA UG theory imply restrictions on the way in which learners learn grammar, although their opinion about the source of these restrictions may differ (the former is functionalist, while the latter is nativist). In this case, the ideal language acquisition intervention, be it explicit instruction or simply the actions taken by an individual learner, would be to expose the learner to certain kinds of linguistic input at particular points, based on a perfect knowledge of the status of the learner's current $L2$ "interlanguage", i.e. the grammatical rules that the learner has so far been able to infer about the $L2$.

The key difficulty here is that this is an AI-hard, or AI-complete problem, one that is beyond current Artificial Intelligence (AI) algorithms (Fass, 2006). Precisely assessing a learner's $L2$ competence is a challenging task for a human language teacher, or even for the learner themselves. It is particularly challenging in that the key SLA observation that all researchers agree on is that there is a difference between what learners can produce in tests versus what they produce in real linguistic situations, i.e. there is some disconnect between implicit and explicit linguistic knowledge. A hypothetical MALL system able to instantaneously assess learner language competence with near perfect accuracy could operate to grammatically simplify something said by a native speaker. This would ensure the learner was receiving input that was beneficial in terms of maximizing their development through the grammatical hierarchy.

Unfortunately no such magical language assessment device exists, so the implication from these theories for MALL appears to be limited, except to caution that expectations about the influence of mobile instruction, or mobile supported language activities, should not be too high since there are aspects of language learning that are constrained, i.e. no amount of instruction or scaffolding will change the time course of some $L2$ developments. However, these grammatical SLA theories are more than simply constraints on language learners; they are exploring the process of language learning and the specific cognitive processes that might be involved. Processability Theory certainly indicates how a language course should be structured, at least in terms of when certain grammatical concepts might be introduced, although this would presumably apply equally to a classroom or a mobile environment. As an example, Cui & Bull's (2005) study delivers grammar exercises to mobile devices, and stores information about past performance, i.e. creates a student model. This is an interesting study, but seems to replicate the classroom instruction model on a mobile device, and the main benefit of mobility is simply convenience.

2.1.2. *Error focused SLA theories*

One important SLA tradition is valuing errors as learning opportunities rather than seeing them as something to eradicate (Corder, 1967). Caroll (2007) suggests that language acquisition is only triggered when there is a parsing/processing failure, and this can only happen during the act of comprehension. Attempting to comprehend input is a key focus of several SLA theories which have roots in the Comprehensible Input hypothesis, one of a number of hypotheses in Krashen's (1981) Monitor Theory. Comprehensible Input suggests that acquisition occurs when learners understand messages that are just beyond their current stage of development. While this hypothesis and other aspects of Monitor theory have been criticized (VanPatten & Williams, 2007) they are still very influential in the SLA field. For example, Input Processing Theory (Van Patten, 2007) explains that a good deal of language acquisition is dependent on learners correctly interpreting what a sentence means, and proposes an instructional intervention called "processing instruction" that seeks to intervene during input processing, thus altering learner's processing behaviors and leading to more grammatically rich intake. For example, learners may rely on lexical items like "yesterday" and "tomorrow" to understand an utterance, without ever processing the grammatical tense markers that indicate past or future. Processing Instruction involves asking learners to distinguish between sentences and utterances that are only differentiated by the grammatical distinction in order to facilitate acquisition. The implication for an ideal MALL system would be to record utterances in context, and dynamically construct counter-examples, e.g. if the learner hears an utterance beyond their current grammatical level and bootstraps a response, then the MALL system could immediately provide paired examples and counter-examples only differentiated in terms of the target linguistic construct. Again, this is an AI-Hard problem.

Another SLA theory with an associated approach to instruction is Gass & Mackay's (2007) Interaction Framework; which incorporates not only the comprehensible input hypothesis but also the comprehensible output hypothesis (Swain, 2005) that producing comprehensible output is essential to acquisition, i.e. learners who fail to speak and write the $L2$ will not achieve native competency. As well as placing importance on the availability of comprehensible input and the generation of comprehensible output, the Interaction Framework focuses on learner interaction, negotiation over meaning and various types of feedback as the key mechanisms for language acquisition. An approach to instruction called Task Based Learning (Ellis, 2003) is informed by the Interaction Framework and advocates instruction around real world tasks such as calling an airline to confirm a reservation. There is clear application to MALL, in that a learner might well be using a mobile device to perform such a real world task. The importance of a task like calling an airline to confirm a reservation is that it provides a backdrop for the negotiation of meaning to take place, i.e. one speaker trying to confirm what the other speaker meant. Rounds of successive feedback allow the speakers to reach agreement on what is meant, and the IF argues that this kind of activity is essential to language learning. The challenge is the role for a mobile intervention. One can imagine a real situation where a non-native speaker tries to hire a car, or book a flight in $L2$, a real situation where they might have their mobile device present. There are at least two possible roles for the mobile device in this situation. One to actually help them achieve their goal, and the other to help them get the greatest language learning benefit. Just struggling through as people do is likely to be beneficial in itself, but may be too challenging or intimidating for many. The role of the mobile device can be to make it more convenient to carry some kind of electronic phrasebook, to record the event for later review by the learner, to try and intervene through some sort of automatic translation system, or to help a learner connect with other learners and/or native speakers.

Thus the most interesting mobile interventions implied by both grammatical and error-driven SLA theories appear difficult to take up without more sophisticated linguistic ability on the part of the mobile device. However, in general terms, it seems that the mobile device should provide, or help the learner encounter, input that is both comprehensible and just above comprehension level; scaffolding the generation of comprehensible output, and above all facilitating linguistic interaction. In particular, given the importance placed on real world tasks, a MALL system is ideally placed to provide some sort of support in a real world situation, outside the language learning classroom; even if that is just the use of mobile dictionary functionality.[1] This could allow an otherwise incomprehensible sign or utterance to be understood, e.g. the description of an artifact in a museum or the explanation by a curator could be comprehended, and allow the learner to test a guess on their part about the nature of the $L2$ by asking a question related to what they have just understood. Specifically, the vocabulary and phrases supplied by the mobile device might allow the learner to guess that the curator had said that the artifact was from a particular time period. This would allow the learner to form a hypothesis about how "belonging to particular time period" was expressed in the $L2$, and test that hypothesis by asking questions using the phraseology that the curator had just used. In the absence of a mobile device, the curator's utterance might have been too far above the learner's comprehension level to allow any useful hypotheses to be formed and thus a valuable learning experience would have been lost.

One older SLA theory with potential implications for MALL is Doughty's (1991) "Cognitive Processing Model" (CPM) which is seen as especially relevant to CALL. CPM suggests that human knowledge is organized into interrelated patterns of schemata and that language comprehension necessarily requires a match between the learner's existing linguistic schemata and the input they are trying to comprehend. According to Doughty (1991), comprehension consists of three stages:

(1) Locating a schema that appears to match linguistic input.
(2) Finding elements in the input that match roles in schema.
(3) Making inferences to cover any gaps that emerge.

The implication is that the process of SLA is difficult because learners may not possess the necessary cultural and linguistic schemata. CPM indicates that comprehensibility is underpinned by the ability to identify relevant schema or contexts. Thus, the mobile device should provide material relevant to the user's context, e.g. by having language materials relevant to an exhibition that the learner may be attending. This is how the mobility construct differentiates itself from classroom learning, as it can act as a learning scaffold in many more contexts than regularly occur in the classroom. To continue our museum example from above, there are vocabulary and phrases that are particularly relevant to the museum context.

[1]Clearly this could be achieved with a compact phrasebook — the advantage of the mobile search is access to a wider range of words and phrases than a phrasebook, the ability to perform search, play audio of word pronunciations, see multiple alternate images expressing the meaning of the word or phrase etc.

Indeed one might argue that the types of things said in a museum context would fall into a particular schema, and if the mobile device was supplying relevant materials the chances of the learner comprehending input and generating comprehensible output would be increased, further increasing the potential for useful learning experiences.

2.1.3. *CALL theories*

The field of Computer Assisted Language Learning (CALL) also brings some highly relevant observations to the MALL domain. The CALL field has identified clear advantages and disadvantages to the use of technological aids in language learning. In one sense, computers are at an advantage: for example, the early CALL literature emphasized the benefits of privacy and individualization (Kenning & Kenning, 1990), taking advantage of the fact that the computer can adapt to the individual learner and mistakes are not publicly exposed. Mobile devices can of course capitalize on this advantage easily by allowing the mobile device to further adapt to the learner's environment when compared with fixed devices such as ordinary PC. For example, location-based services can adapt not just to the individual, but the situation the individual is in, such as providing food related vocabulary when in a restaurant.

In another sense, the use of computers can be disadvantageous in learning, such as with the challenge of how to handle error tolerance. For example, in a real-life classroom, teachers may tolerate an error in one student to focus on another they think is more important, but correct it for another student. However, this is a far more difficult judgment for a computer to make (Beatty, 2003), although this is not to say that some model of student competence is not useful, just challenging to implement successfully. Furthermore, learners may learn more about the rewards mechanism of a computer program than the material it is trying to teach (Beatty, 2003). A related issue for mobile devices is that users may perceive mobile applications as simply an amusing technological pass-time and thus fail to genuinely engage with the actual content and then be unable to generalize what they learnt from the mobile device to other contexts.

2.1.4. *Theory in implemented MALL systems*

Different learning theories have naturally been reflected in the implementation of existing mobile language learning systems. Thornton & Houser (2005) compared the use of pull (web-based) and push (email) approaches in delivering vocabulary content to mobile phones. Comparison between pre- and post-test on two weekly cycles indicated that subjects more than doubled the amount of vocabulary retained when using the push approach. While Thornton & Houser's study is of considerable interest in terms of the practicality of the different push/pull technologies, their system was clearly based on largely behaviourist principles in as much as small amounts of material were incrementally made available to students with little opportunity for knowledge construction or interaction between learners.

Arguably these kinds of systems do provide users with input that is comprehensible or just above the level of comprehensibility, but they do not utilize location or contextual information in such a way that might help learners identify relevant cultural or linguistic schemata. They are simply attempts to provide content that can be used for learning, to fill in the gaps between other activities.

On the other hand, more constructivist-style approaches have successfully used location information in the learning process; for example, Ogata & Yano (2004) tested their CLUE system by having Japanese native speakers converse with English speakers at a number of locations such as a restaurant, hotel etc. The CLUE system provided helpful words and phrases based on the location of the learner, as well as identifying other learners who had already encountered that material (Figure 1). The learners showed substantial increases in vocabulary recall compared to control on pre- and post-tests. While there may be ethical issues with having learner data automatically shared with others this was an exciting example of the way MALL can operate to take advantage of the mobile construct in a way that could not happen in the classroom and scaffold various type of interaction. See Ogata *et al.* (2008) for further similar studies.

PALLAS (Petersen & Markiewicz, this issue) is a mobile system based on constructivist principles such as situated learning. Like Ogata & Yano's (2004) CLUE system, PALLAS provides vocabulary related to the users current situation, but also personalizes it along a number of other dimensions such as experience, interests, gender, learning ability etc. PALLAS supports both study exercises, and provision of useful vocabulary to help scaffold interaction with other learners and native speakers, but has yet to undergo a learning outcome study such as that conducted on CLUE.

Commercial MALL systems tend to focus on behaviorist-style programmed instruction (Leucker & Ash, 2003), but occasionally one does see support for constructivist elements such as communication and interaction between learners. For example the AppliLearning system makes it possible not only for teachers, but also for learners to provide text based study materials available for review on their cell phones. The material (generally vocabulary) is organized into short study lessons, but learners can also make lessons available to other users in the system. This process of sharing learning materials and making contact with other learners can be seen as scaffolding constructivist learning activities. However, the authors are not aware of any truly collaborative mobile language learning systems that parallel collaborative language learning web sites like JGram (http://www.jgram.org/).

In support of the constructivist position, various research has shown that if a learner performs deeper processing of material (Groot, 2000), by embedding it into their existing knowledge framework, they are likely to retain it for longer (Chi & Koeske, 1983); particularly to the extent that it is rehearsed through multiple modalities (Chun & Plass, 1996; Gathercole & Conway, 1988). The advantages

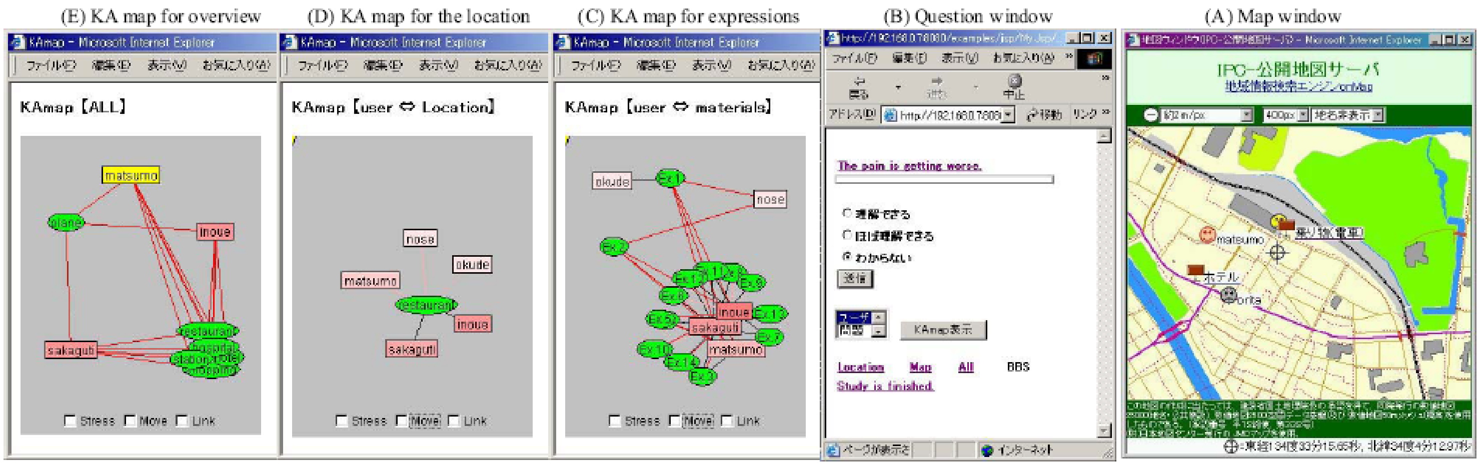

Figure 1. Screenshots from the CLUE system (Ogata & Yano, 2004) showing from left to right, three vocabulary concept maps: a quiz; location of materials on map of “town”.

of rehearsal in multiple modalities is explained by Dual Coding Theory (Paivio, Clark & Lambert, 1988) which suggests that visual and verbal information are processed along separate pathways in the brain, leading to greater connectivity and increased chance of recall when both pathways are involved in learning. This concept is elaborated upon in the application of Multimedia Learning Theory to SLA (Plass & Jones, 2005). As well as reviewing the evidence for dual coding theory Plass integrates multimedia learning theory with the Interaction Framework, generating two further principles: Individual differences and Advance Organizer. The individual differences principle reflects the finding that different learners have different styles, i.e. they are visualizers or verbalizers, and presenting learners with both visual and verbal modes and allowing them to choose their focus is more effective, and allows learners to manage their cognitive load. The Advance Organizer principle stems from research showing that language learners gain more from videos when they are exposed to short relevant summary materials (e.g. select words and phrases) in advance.

Thus there are several key implications for MALL from the various theories presented above:

(1) Present material at the level, or just beyond the level, of the learners current ability.
(2) Create authentic task based learning.
(3) Scaffold interaction with others.
(4) Connect with learner's existing knowledge schemas.
(5) Present both visual and verbal information in tandem.
(6) Allow learners the choice of modality.
(7) Give learners advance preparation.

All of these points apply just as well to the language learning classroom, or CALL environment. However, each has different interactions with the mobile construct. The capacity to provide material at the correct level, or in a particular modality is constrained by the nature of the mobile device (as we shall see in the next section). The MALL approach potentially excels in scaffolding task based learning in the real world, and connecting learners with other learners and native speakers.

3. Support for Visual Media

As we discussed in the previous section, support for additional media types on mobile devices is likely to be beneficial to language learning to the extent that the modality can be selected by the learner, and that it encourages deeper processing, provides opportunities for the learner to map input onto their own relevant schema. This should allow the learner to make adjustments to their existing schema, and even create new ones. Unfortunately, provision of multi-media content on mobile devices can be challenging because audio and graphics file sizes are generally much larger than the default text content, a factor to consider carefully given the restricted storage space of most mobile devices.

Pure text-based study requires either support for the native language of the learner, or operates exclusively in the target language, thereby requiring a careful match between content and learner competency in the target language. However, even if the native language of the learner can be supported, concept mismatch between target and native language may still cause problems. For example, words, phrases and idioms in one language do not necessarily map directly onto those in another language, making it challenging to provide a single correct translation. Although visual media cannot completely resolve this problem (which involves issues of reference and interpretation), it provides a different angle of attack. Learners will often have formed concepts by generalizing over a set of visual experiences (Eimas & Quinn, 1994), and thus visual materials can replicate some of this process, e.g. a child learns the word "purple" after hearing the word "purple" in multiple contexts when certain colors have been visible. Given suitably equipped mobile devices such as camera-phones, it is relatively straightforward to capture images and videos of different events and artifacts which can then be re-presented to native speakers to assist vocabulary acquisition (e.g. "what do you call this?"), and even as a sort of picture dictionary to assist in conversational settings.

Image and video capture on mobile devices is of course limited to the resolution of the camera involved and the amount of storage available on the mobile device. The first commercial cameraphone had a resolution of 0.11 megapixels, and 200 Kb storage: roughly equivalent to the capacity to store about 40 jpeg images. Camera phone resolution is now in the 10 megapixel range, and with storage now in the gigabyte range, language learners are unlikely to be limited unless they are working with more demanding video content.

While the individual learner may be content to create a picture dictionary on their phone, a more constructivist approach would suggest that learning will be more effective if learners are given the opportunity to collaborate. Of course, regardless of whether for collaboration, or simply to back up image data, the bandwidth of data transfer between device and server becomes the more serious bottleneck for mobile devices. Different image formats can provide relief from this bottleneck by supporting greater compression; however the degree of image compression provided by a particular format is often dependent on the nature of the image being compressed. For example, the JPEG format can typically achieve 10:1 to 20:1 compression without visible loss, and is generally considered superior to the PNG format, but these ratios are not consistent over all image types (Miano, 1999). In addition, moving data to the other devices is naturally restricted by the data transfer mechanisms available to the mobile device, namely:

- Physical link, e.g. USB: fast but only possible when there is a fixed device to connect with. Also, the requirement to make an actual physical connection can be inconvenient.

- Short range wireless link, e.g. Bluetooth: fast but needs to be close to a bluetooth source (∼10 m).
- Medium range wireless link, e.g. WiFi: fast but need to within relatively small radius of transmitter (∼30 m).
- Long range wireless link, such as radio (e.g. GSM, GPRS or CDMA) or WiMax: Radio links have low bandwidth, but new technologies such as WiMax and EVDO are increasing available bandwidth.

Naturally, radio-based cellular networks provided by phone companies will have charges associated with them, making many users reluctant to use them. It is also worthy to note that even using advanced "3G" networks it can take 10–20 seconds to transfer the 400–500 Kb image generated by a 2 megapixel camera phone, indicating that it may be a little while before we see large communities of language learners sharing photo flashcards with each other.

Despite these limitations, PhotoStudy (Joseph *et al.*, 2005) took a first step towards creating such a community by supporting vocabulary study on both wired and wireless devices. PhotoStudy allowed users to email photo attachments from fixed or wireless devices with vocabulary terms in the subject line. These images were then entered into a central database and linked with their associated vocabulary terms.

Users could then download lessons comprised of sets of images where they attempted to guess the correct vocabulary term associated with an image. Figure 2 shows an example of study screens within an example lesson. The big advantage of PhotoStudy as a method of vocabulary acquisition is that it allowed the creation of a set of photo flash cards tailored to the users own learning environment, e.g. photos of actual artifacts encountered by the user can be employed, connecting vocabulary study to the broader context in which the learner exists. Arguably if learners can connect vocabulary terms to artifacts relevant to their everyday life (artifacts that are themselves embedded in the learners existing knowledge web), then the learners will have a greater chance of being able to access the cultural and linguistic schemata necessary to achieve comprehension in different learning contexts.

The key contribution of the mobile device is reducing the barrier the learner has to generating study content out of the things they encounter in their environment. Clearly the same effect could be achieved with a digital camera and desktop computer, but making the effort required to generate new study content absolutely minimal, maximizes the chance the learner will continue to generate new content over time.

PhotoStudy also provided some support for collaboration by allowing users to create quizzes out of certain subsets of images and sending them to friends. There is a need for "policing" of the study material, since even the simple PhotoStudy prototype became victim of email spam that led to failed image links. There are also issues of ambiguity concerning which part of the image the vocabulary term is

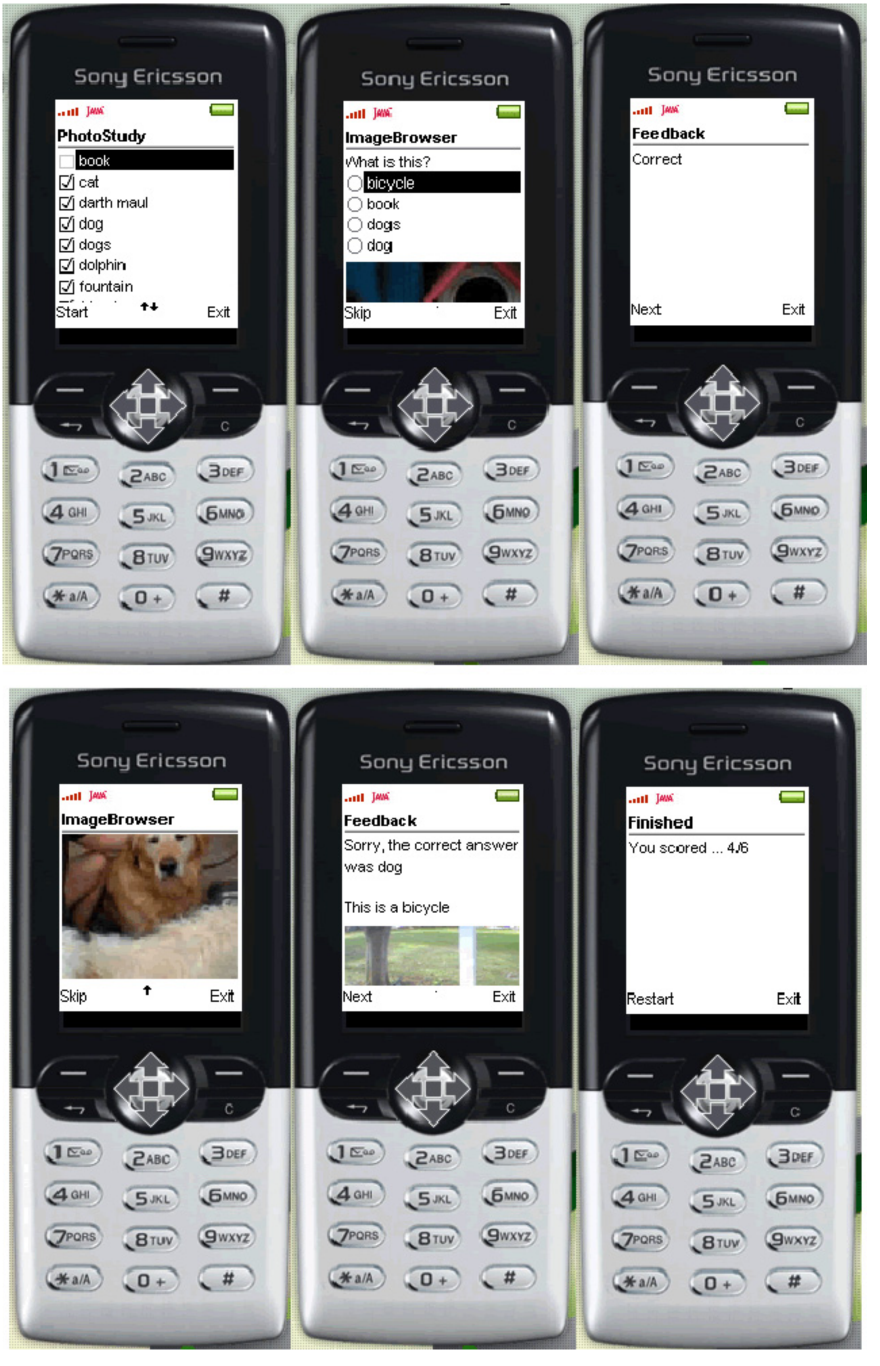

Figure 2. Taking a photo flashcard quiz with PhotoStudy.

referring to, and differences of opinion about whether a particular image really is representative of a given term.

Of course, contributed content could be moderated, but this places a burden on a hypothetical moderator. Storing the frequencies with which different study materials are referenced and shared could support some form of "popularity" navigation. Additional care is needed to create a mobile interface that supports navigation of a large amount of material. The key should be to allow ordering of material by category, age, popularity and so forth; allowing users to set default orderings and providing a single link to sets of other options. Of course popular content could be the result of frequent misperceptions, so there is certainly a role for a moderator or native-speaking participants to provide checks, i.e. to look at how an image has been labeled by a non-native speaker and suggest alternatives. Ideally more than one language would be supported, and native speakers of different languages could provide adjudication support for each other, e.g. "I'll help you correct your English photo flashcards, if you help me with my Japanese ones".

Other MALL systems that make use of visual content have often been PDA based (Thornton & Houser, 2003; Garcia Cabrere, 2002). PDAs have other advantages compared to phones. They afford for larger screens and higher resolutions, naturally making the visual experience more rewarding, although many users may prefer not to carry an extra device in addition to the mobile phone that they were already carrying. A point to consider in this respect is how familiar learners are with mobile phones compared to PDAs. According to the Gartner Research, 15 million PDAs were sold in 2005, compared with 816 million mobile phones, making it clear that the mobile phone is a more ubiquitous device. Nonetheless, the PDA may still be a good alternative for particular situations, such as classroom exercises.

4. Support for Auditory Media

The auditory modality can also provide substantial benefit to the learner in several key areas which are critical for successful language acquisition. Especially important for the language student is learning the skills to accurately generate comprehensible output, as well as to correctly perceive what is being said to them. Hence scaffolding correct pronunciation and perception is essential for developing successful oral/aural language proficiency. Fundamental to mastering these skills is the use of auditory media, which could be used in mobile device applications in some of the following ways.

4.1. *Pronunciation guide*

It is important for language learners to know how to pronounce a new word that they are not familiar with. This allows them to generate comprehensible output, which is necessary for second language acquisition. CALL applications can provide obvious assistance in this respect, as they can provide a single (or multiple) "model" or "ideal" speech samples for the learner to listen to.

4.2. *Pronunciation correction*

There is potential for CALL applications to sample input from the learner and (particularly if used in conjunction with Automatic Speech Recognition or ASR engines) determine whether the pronunciation of key sounds (such as vowels or difficult consonants) are articulated in a near native-like manner. There have already been some initial attempts at this kind of application in the mobile device arena. One study in this area for example (Yang *et al.*, 2005), scored the "correctness" of an answer to questions based on input into the mobile device. Thus, the learner would work on their pronunciation in order to answer the questions correctly.

4.3. *Pronunciation modeling*

CALL applications could also potentially offer modeling of the spoken input as an average over time. This could determine patterns or common errors made. Apart from being useful at an individual level, there is also potential for data from a number of language learners of specific language groups to be pooled, allowing the instructors and course designers to focus their efforts to correct known problems that would pose issues for specific language groups. For example, in the case of Japanese speakers, it has been found that the /r/–/l/ contrast in English is difficult because of a failure to distinguish (and therefore articulate) in the third formant boundary (see Uther *et al.*, 2005a,b for a review). Of course there are other notable examples of $L2$ pronunciation difficulties that are well established (e.g. Spanish pronunciation of English vowels as one example). However, even within language groups, there is a fair amount of individual variation in $L2$ phoneme difficulties, so the ability to personalize the learning application to an individual would be very useful. However, to our knowledge, this kind of pronunciation modeling at an individual level has not yet been attempted.

4.4. *Perceptual/phonemic discrimination problems*

CALL applications in the mobile domain can also easily be adapted to correct perceptual/phonemic discrimination problems that cause specific problems in oral and aural fluency for the language learner. There are some good techniques which can (and have been) easily adapted to the mobile environment to allow the delivery of training interventions to focus on these type of specific phonetic category problems. Interventions of this kind have been presented in previous work (Uther *et al.*, 2005c) and have been shown to have excellent learning outcomes (Uther *et al.*, 2007). These are reviewed in detail later in this section.

4.5. *Conversational practice to assist fluency*

Finally, one perhaps somewhat overlooked use of mobile technologies is the use of mobile devices to assist conversational practice. This could be with either a real or virtual tutor. Mobile devices (particularly phones) are very much suited for this kind of interaction. They have a natural affordance for audio interaction as we primarily use phones to speak into and listen from. Therefore, there are excellent reasons for exploring this aspect for mobile language learning.

4.6. *Technical issues relating to audio MALL*

When developing CALL applications, it is also useful to consider that there are specific technical issues which may constrain the use of audio capability (either input or output). Certainly, this has been the experience of one of the mobile language learning applications (MAC) that we draw experience from. The MAC software built directly on the work done using a traditional PC environment and focused on the problem of Japanese /r/–/l/ discrimination, which is notoriously difficult for Japanese language learners of English, but very amenable to training with PC-based training using a simple application (called "high-variability training") that gives learners practice and feedback on a forced choice discrimination task on the /r/–/l/ contrast (Logan *et al.*, 1991; Lively *et al.*, 1993; Bradlow *et al.*, 1994, 1997, 1999). Within the application, the learner is presented with two words on a screen and then presented with a spoken word. The learner's task is then to respond correctly with the written word that corresponds to the one they heard. The name "high variability" comes from the fact that the application will vary phonetic contexts of the words (i.e. the target contrast may be either at the start, such as "rake" vs "lake" or else the target contrast may be in the middle, such as "miller" vs "mirror"). MAC was a novel variant of the high variability method firstly by its platform — a mobile phone. Secondly, it also trialed a new technique that used adaptive algorithms to tailor the training for the learners. The MAC software adapts according to the learner's responses and presents to the learner a speech contrast of the type in which they will most need further practice (e.g. if they tended to make more mistakes in initial position contrast words, they got greater practice in those type of words).

Whilst MAC did not necessarily use the mobile context in its implementation and execution, the device nonetheless was a good example of how a personal trusted device can use personalized profiles of the learner in its training programs. Although this theoretically could also be accomplished using a PC, the mobility was clearly far more convenient for the learner. Additionally, further iterations of the application have also now been applied to Finnish speakers of English and their perception of vowels (Ylinen *et al.*, submitted). Interesting future possibilities that take more advantage of the mobile context include allowing learners to record audio samples of native speech to create their own "audio flash cards".

In delivering the MAC software, one of the first and foremost technical issues that needed to be addressed related to the sound quality. Of course in an application like MAC, it is critical to have a good sound quality as the application needs not only a clear distinction between the speech sounds but also a realistic and natural sounding quality of voice, to ensure generalization to the real-world context. Sound quality is in turn affected by two other main factors: the memory size (as sound files are often large) and the type of audio encoding algorithm used. There are also other factors such as hardware (e.g. speaker quality), but in most mobile phones, these tend to be generally of high enough quality so as to render voice easily intelligible.

Regarding sound file sizes, it is useful to consider the options for audio encoding in mobile devices, which may allow compression of some of signal and can vary enormously in their quality. Audio encoding algorithms generally may be divided into those that preserve all features of the original waveform (lossless) and those that discard features in order to aid compression (lossy). For mobile phones, where storage and bandwidth are at a premium, lossless codecs such as WAV are rarely used (although in PDAs, one can use WAV, these samples will be much larger). However, lossy codecs such as MPEG layer 3 (MP3) and Advanced Audio Codec (AAC) are able to reduce the file size of most audio content by a factor of 10 with almost no audible sound quality loss. A comprehensive review of audio codecs used by mobile devices may be found in Autti & Biström (2004). Although most phones that support Java applications can play WAV and AMR sound samples, their support of other codecs (such as MP3) is patchy.

As speech applications may easily use a corpus of several hundred speech samples, file size consideration is paramount. For some common mobile phones, it would not be unusual to have only 10 Mb of memory available for application use. It is tempting therefore to simply opt for the lower file size codecs, but there is a trade-off in sound quality (and therefore training quality). Some early attempts at using speech samples in our MAC pilots with codecs at the smaller end of the file size spectrum (AMR) showed that the sound quality was not sufficient enough to deliver good quality training (Uther *et al.*, 2005a). More recent investigations of the capabilities of newer hardware has shown that support for MP3 files from Java applications is now much more common, and that the memory capacity is large enough to support a corpus of speech samples in MP3. Subsequent implementations of speech-based CALL applications using MP3 format for superior sound quality resulted in a statistically significant improvement of about 15% from baseline (Uther *et al.*, 2005a, 2007). MP3 seems a good compromise for the design of audio-based mobile apps as it offers substantial file size reductions, with virtually no noticeable trade off in audio quality.

Another technical consideration for the designer of audio-rich mobile applications is the potential for using speech samples collected by the device itself and using an ASR engine to analyze the input for accuracy of pronunciation. Of course, for CALL applications to successfully offer good quality speech recognition using ASR technology, it must have decent quality input. Quality input in turn depends on the quality of the microphone, a good frequency response and sampling rate. Despite the fact that mobile phone hardware in particular is generally optimised for voice, these parameters may not always be optimal for ASR engines and will vary from manufacturer to manufacturer and model to model. It is therefore critical when designing applications that use ASR engines to think of the constraints that specific hardware may impose and the consequences this may have on the kind of input received.

So far we have mainly focused attention on the phone and PDA use. Of course, a newer type of mobile device is the iPod (or generically, the portable mp3 player). These devices can and have been used in CALL applications. For example, a trial at Duke University was used to support Spanish and Turkish language learning (amongst other trials). Students in a Spanish class used iPods to respond to verbal quizzes, submit audio assignments, record audio journals, and receive oral feedback from their instructor. A Turkish class used them to listen to authentic materials such as news, songs, and poems, and to the instructor's vocabulary and translations (Belanger, 2005).

Other kinds of mobile devices that support audio capability include digital voice recorders and mini camcorders. These can also be used to support language learning. For example, students taking distance-learning German and Spanish courses through the United Kingdom's Open University used digital voice recorders and mini-camcorders to record interviews with other students and locals and to create audiovisual tours (Kukulska-Hulme, 2005).

5. Future Developments

Constructivist theory suggests that "situated learning" is one of the most effective means of achieving educational goals. While classroom study is important, more often than not the things being taught in the classroom relate to other things occurring elsewhere in the world, as indicated by the Interaction Framework informed task based approach to learning. For example, in a language classroom a learner may engage in a role-play where they pretend to be shopping for food. This is a representation of the real activity of shopping for food at a store, and thus the role-play is only as effective as to the extent to which it can replicate the actual activity. Clearly there is the additional advantage that the classroom provides a safe environment for the learner, but at some point the learner needs to practice their skills outside the classroom, so that they can generalize what they have learnt into other contexts. A mobile learning system can provide the learner with scaffolding outside the classroom, by allowing them to store and record media that reflects their actual experience. Many classes use media such as textbooks and audio-visual aids that may or may not have special meaning to a particular learner. A mobile learning system can allow each individual learner to prepare and share content that has personal meaning, helping them individualize and reinforce the learning process. Arguably the time spent on these sorts of activities could be spent on depersonalized drills that take place out of context; however, it seems clear that while this may boost explicit linguistic knowledge, it fails to provide the necessary implicit linguistic knowledge to cope with real-life situations. Motivation is also a key issue, and individualized content that has personal meaning is likely to boost motivation more that repetitive drilling (Dornyei, 2001a).

A truly mobile learning system could allow individual learners to tag items in their environment by capturing them in some media form. A simple example would be to take a picture of the local post-office, and store it with GPS location data. The mobile language learning system would associate a tag for this data which could be the target language word for post-office. The learner would subsequently be able to access their media through a map view, or when they returned to the location they originally captured it at. Depending on their degree of altruism, other learners could discover their data through arriving at the same place. This kind of framework would offer excellent potential for learners to share materials by "leaving" or tagging them with their current GPS location for other learners to subsequently encounter when they approach the same location (Rashid *et al.*, 2006).

Immersive language learning systems that make use of 3D environments already exist (Johnson *et al.*, 2005; Purushotma, 2005), but currently they are disconnected from any physical location. These sorts of systems can be thought of as a kind of Virtual Reality (VR) a concept that has become familiar to most. However the related concept of Augmented Reality (AR; Azuma, 1997) is perhaps less well understood. A VR system is generally thought of as a computer simulation of a real or imaginary system that enables a user to perform operations on the simulated system and see the effects in real time. While users will consciously enter a VR system, AR tends to imply that interaction with an information system will be unconscious and potentially unintentional, i.e. the artificial elements of the augmented reality will become effectively indistinguishable from the actual reality — the goal is to "augment" or supplement reality, not to replace it. Many AR projects call for the user to wear a pair of glasses that allows them to see an additional reality layered over the existing one.

We can think of this as a more sophisticated version of the location based mobile learning concept described above. In an MALL system, a device carried by the user might vibrate to alert the user to the existence of some tagged information in their immediate vicinity. Using an AR system built into a pair of eyeglasses the user can be see things such as virtual post-it notes and graffiti in real-world locations merely by turning and scanning their environment. An interesting study presented by Beaudin *et al.* (2007) involves the use of a special house where each item is labeled with an RFID tag. When the learner comes close to an item, they hear audio of that word in $L1$ and $L2$. This is a fascinating study into the possibilities of context sensitive learning of vocabulary.

Support for audio and visual media have been heavily developed on mobile devices, to the extent that many cellphones now support streaming video and polyphonic sound. It seems likely that these will be the forerunners of mobile AR. In addition, haptic (Chang & Sullivan, 2005), olfactory (Yanagida *et al.*, 2004) and gustatory (Bach-y-Rita *et al.*, 2003) interfaces have been developed, and in some cases, integrated into existing mobile phones. Theoretically multi-modal stimulation will likely help learners embed concepts in their existing web of knowledge, but it may be some time before such technology using other senses becomes widespread enough to be of use in language learning contexts. For example, in language learning environments, certain smells can be paired with meaning in the new language to allow for a richer and more memorable learning experience.

Of course this does not stop us speculating about such systems, which may provide experiences of the vocabulary users are learning; haptic experiences for the terms "rough" and "smooth", olfactory experiences for "malodorous" and "perfumed", and gustatory experiences for "bitter" and "sweet". Arguably with such systems we are placing the learner in situations much closer to their experience of learning their native language, where the meaning of different terms evolved from sensory experiences coupled with spoken and written forms. Naturally there is a lot more to language than learning words by association with experiences, and such futuristic interface modalities may not help scaffold the social interaction that is really necessary for learning a new language. However the ability to share virtual experiences through multiple senses could act as an incentive for people to interact socially.

Finally, from a technical perspective, there are also a number of devices that are beginning to support the use of programming languages and environments that are not only much easier to use, but far more platform independent. In the past, programmers have had to rely on programming in platform-specific environments such as Symbian or Java. Although the latter is theoretically platform independent, our experience is that the portability of an application from one mobile phone to another is not at all straightforward. This appears to be due to bugs in the Java implementations and fundamental differences in hardware capabilities. However, to ameliorate this problem, technologies such as Flash are now beginning to be supported in mobile phones. Such environments promise to be easier to use and offer easier portability of applications between devices. If this indeed proves to be the case, this will of course pave the way for the development of many new applications for mobile devices, and thereby allow them to be more easily used for language learning applications.

6. Evaluation and Testing

There are certain technical issues with evaluating and testing all mobile systems, that must be considered when developing MALL applications. One issue is that it is more challenging to collect data from a mobile device compared to a PC. In the MAC system, getting data to the actual phone was complicated by the fact that if the session ended prematurely or the user did not prompt the machine to write to the memory card, data would be lost. In contrast, retrieving usage data from PCs tends to be more reliable and certainly easier to implement.

As well as issues with reliability there are difficulties associated in retrieving good usability data from mobile devices. Arhippainen and Tähti (2003) provide some good examples of the difficulty of video recording mobile participants. If a participant is sat in front of a PC, there are many software packages available that will capture a video of their expression along with a time-indexed video of their activities on the PC screen, whereas none appear to exist for mobile devices. Thus on the PhotoStudy system a separate layer of UI recording software was incorporated to create a log of keypresses on the part of the user that could be played back to create a simulation of what the user had experienced on the device.

However one issue that needs to be controlled for in the mobile context is that there are more opportunities for practice and *ad hoc* experience with mobile applications than there are for a PC. Any advantages in the use of a mobile system could be explained by mere exposure effects, thus making reliable data logging crucial. One possible way round this is to ask participants to use mobile simulators on PCs. However in the MAC system for example, there were differences when data was collected from UI studies run on a simulator as compared to an actual mobile device. For example, we saw users make certain errors with the simulator that they did not make with the phone implementation and vice versa (Uther *et al.*, 2005a). Running studies on a simulator was clearly important, but it does not generalize perfectly to use of the mobile device.

In terms of evaluating learning outcomes from the use of mobile language learning applications, there are of course several formal methods which can be used to achieve the different kinds of evaluation. Unfortunately there is no definitive answer as to which is the best approach to take, since it depends largely on the sample being evaluated, the scope of the evaluation and even the intervention itself.

One of the most rigorous and objective ways of testing the efficacy of an intervention is with an experimental design. For assessing a specific intervention, the within groups (repeated measures) design is usually appropriate. This would involve measuring the students' performance at pre- and post-intervention stages (i.e. before or after the course). Specifically for language learning interventions, the variables which one would use to assess the outcomes would fall into the following key areas: written, oral and aural examination.

For written proficiency, any number of tests may be designed to assess written comprehension and fluency. Some examples include tasks such as reading short passages and answering comprehension questions. Another possibility is to require a verbal word to be transcribed and spelt correctly.

For oral proficiency, it is useful to collect simple production samples of the new words the learners are learning. These samples can be later independently rated either by teachers or independent native-speaker raters for the degree to which they are close to native-like pronunciation. There is also a theoretical possibility that speech samples could be scored automatically using speech recognition systems. General oral fluency may be also assessed using conversational and reading tests as well.

For aural proficiency, again there are many possible tasks that could be used. One common test would be aural comprehension or conversational tests. Another often overlooked possibility is perceptual/phonetic discrimination and categorization tasks; useful in specific cases where one language group finds that due to the phonology of their first language (or $L1$) that this causes problems for learning a second language (or $L2$).

7. Conclusion

The use of mobile devices and their multimedia capabilities can help language learners have more authentic real world learning experiences, situating learning within their cultural and linguistic schemata. Whilst there are undoubtedly technical obstacles that currently limit multimedia on mobile devices, these are gradually disappearing. The value of using mobile devices and incorporating multimedia elements into language learning applications needs to be quantified with controlled experiments where control groups study on non-mobile platforms, or in mobile contexts with non-technical support, e.g. paper flashcards. These sorts of experiments should be a priority for future research.

The initial results of various attempts to use multi-media in support of language learning on mobile devices look promising, and there are many recommendations to be drawn from SLA and CALL theories. All of these have been tested or evaluated in some form, but large user studies with controls are necessary to tease apart the relative benefits of each. To re-iterate, those recommendations are:

(1) Present material at the level, or just beyond the level, of the learners current ability.
(2) Create authentic task based learning.
(3) Scaffold interaction with others.
(4) Connect with learner's existing knowledge schemas.
(5) Present both visual and verbal information in tandem.
(6) Allow learners the choice of modality.
(7) Give learners advance preparation.

In this paper, we have argued that mobile user generated media such as photos or audio recordings can support learners in connecting words and phrases to episodes relevant to their everyday life. This in turn gives learners a greater chance of being able to access the cultural and linguistic schemata necessary to achieve comprehension in different learning contexts. The process of generating multimedia learning materials itself is a useful instance of task based learning. Furthermore it is important for the language student to accurately generate comprehensible output, as well as to correctly comprehend linguistic input. Hence, scaffolding correct pronunciation and perception is essential for developing successful oral/aural language proficiency. While this does not require explicit use of the mobile context, the convenient microlearning opportunities provided by the mobile platform are a considerable advantage. Thus whether through the mobile context, or simply the convenience of the mobile platform, we have argued that there are benefits to be gained from multimedia MALL, and that careful experimentation is needed to identify the optimal balance of approaches.

Acknowledgments

The authors thank Dr. James Uther, of University of Sydney for his helpful comments on sections of this paper.

References

Arhippainen, L., & Tähti, M. (2003). Empirical evaluation of user experience in two adaptive prototypes. In *Proc. 2nd international conference on mobile and ubiquitous multimedia*. Norrköping, Sweden.

Autti, H., & Biström, J. (2004). Mobile Audio — from MP3 to AAC and further, Multimedia Seminar. Telecom. Software & Multimedia Laboratory. Helsinki University of Technology. Finland, 2004.

Azuma, R. T. (1997). A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments *6*, 4 (August 1997), 355–385.

Bach-y-Rita, P., Kaczmarek, K. A., & Tyler, M. E. (2003). A tongue-based tactile display for portrayal of environmental characteristics. In L. J. Hettlinger & M. W. Haas, (Eds.), *Virtual and adaptive environments* (pp. 169–186). Mahwah, NJ: Erlbaum.

Bailin, A. (1995). AI and language learning: Theory and evaluations. In V. M. Holland, J. D. Kaplan & M. R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Beatty, K. (2003). *Teaching and researching: Computer-assisted language learning*. Longman Press.

Beaudin, J. S., Intille, S. S., Munguia Tapia, E., Rockinson, R., & Morris, M. E. (2007). Context-sensitive microlearning of foreign language vocabulary on a mobile device. In *Proc. European ambient intelligence conference 2007*.

Belanger, Y. (2005, June). Duke University iPod first year experience final evaluation report. Retrieved June 24, 2005, from http://cit.duke.edu/pdf/ipod_initiative_04_05.pdf

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., Tokhura, Y., Lively, S. E., & Yamada, T. (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.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tokhura, Y. (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*(10), 2299–2310.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tokhura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long term retention of learning in perception and production. *Percept Psychophsy.*, *61*(5), 977–985.

Carroll, S. (2006). The autonomous induction theory. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 155–173). Erlbaum.

Chang, A., & O'Sullivan, C. (2005). Audio-haptic feedback in mobile phones. CHI'05, 1264–1267.

Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, *19*(1), 29–39.

Chinnery, G. M. (2006). Going to the MALL: Mobile Assisted Language Learning. *Language Learning & Technology*, *10*(1), 9–16.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*, 183–197.

Corder, S. P. (1967). The significance of learner's errors. *IRAL*, *5*, 161–170.

Cui, Y., & Bull, S. (2005). Context and learner modelling for the mobile foreign language learner, *System*, *33*(2), 353–367.

Dörnyei, Z. (2001a). New themes and approaches in second language motivation research. *Annual Review of Applied Linguistics*, *21*, 43–59.

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition*, *13*, 431–469.

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903–917.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2003). Task-based language learning and teaching. *Modern Language Journal*, *89*(2), 297–298.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. MIT Press.

Fass, L. F. (2006). AI and cognitive science can learn alot from "language learning". In *Proc. of Computational Intelligence*.

Garcia Cabrere, J. C. (2002). Tele-enREDando.com: A multimedia WEB-CALL software for mobile phones. *International Journal of English Studies*, *2*(1), 167–178.

Gass, S. M., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 175–200). Lawrence Erlbaum.

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Mem. Cogn.*, *16*, 110–119.

Groot, P. J. M. (2000). Computer assisted second language vocabulary learning. *Language Learning & Technology*, *4*(1), 60–81.

Johnson, W. L., Vilhjálmsson, H. H., & Marsella, S. (2005). Serious games for language learning: How much game, how much AI? In AIED.

Joseph, S. R. H., Binsted, K., & Suthers, D. (2005). PhotoStudy: Vocabulary Learning and Collaboration on Fixed & Mobile Devices. WMTE'05.

Kadyte, V. (2003). Learning can happen anywhere: A mobile system for language. MLEARN'03.

Kenning, M.-M., & Kenning, M. J. (1990). *Computers and language learning: Current theory and practice*. New York: Ellis Horwood.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Oxford University Press.

Kukulska-Hulme, A. (2005, May 12). The mobile language learner: Now and in the future. Fran Vision till Praktik. Language Learning Symposium conducted at Umea University in Sweden. Retrieved July 28, 2005, from http://www2.humlab.umu.se/symposium 2005/program.htm

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL* , *20*(3), 271–289.

Levy, M. (1997). *Computer assisted language learning: Context and conceptualization*. Oxford: Oxford University Press.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (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), 1242–1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*(2), 874–886.

Luecker, W., & Ash, C. (2003). GCSE revision with mobile phones: Developing a Java-based quiz game. MLEARN 2003.

Miano, J. (1999). *Compressed image file formats: JPEG, PNG, GIF, XBM, BMP.* Boston: Addison-Wesley Professional.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Oxford: Oxford University Press.

Ogata, H., & Yano, Y. (2004). Knowledge awareness map for computer-supported ubiquitous language-learning. WMTE'04, pp. 19–26.

Ogata, H., Hui, G. L., Yin, C., Ueda, T., Oishi, Y., & Yano, Y. (2008). LOCH: Supporting mobile language learning outside classrooms. *International Journal of Mobile Learning and Organisation*, *2*(3), 271–282.

Paivio, A., Clark, J., & Lambert, W. E. (1988). Bilingual dual coding theory and semantic repetition effects on recall. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *14*(1), 163–172.

Petersen, S. A., & Markiewicz, J. K. (this issue). Personalized and contextualized language learning: Choose when, where and what. *Research and Practice in Technology-Enhanced Learning*, *4*(1).

Pica, T. (1998). Second language learning through interaction: Multiple perspectives. In Regan, V. (Ed.) *Contemporary approaches to second language acquisition in social context* (pp. 9–31). Dublin: University College Dublin Press.

Pienemann, M. (2007). Processability theory. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 137–154). Lawrence Erlbaum.

Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467–488). New York: Cambridge University Press.

Purushotma (2005). Commentary: You're not studying, you're just... *Language Learning and Technology*, *9*(1), 80–96.

Rashid, O., Mullins, I., Coulton, P., & Edwards, R. (2006). Extending cyberspace: Location based games using cellular phones, *ACM Comp. in Ent.*, *4*(1).

Saville-Troike, M. (2006). *Introducing second language acquisition.* Cambridge, U.K: Cambridge University Press.

Schcolnik, M., Kol, S., & Abarbanel, J. (2006). Constructivism in theory and in practice. *English Teaching Forum*, *44*(4), 12–21.

Stanford Learning Lab (2001). Mobile learning explorations at the Stanford learning lab. *Speaking of Computers*, *55*.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 471–484). Mahwah, NJ: Lawrence Erlbaum.

Thornton, P., & Houser, C. (2003). Using mobile web and video phones in English language teaching: Projects with Japanese college students. In B. Morrison, C. Green & G. Motteram (Eds.), *Directions in CALL: Experience, experiments & evaluation* (pp. 207–224). Hong Kong: English Language Centre, Hong Kong Polytechnic University.

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning*, *21*(3), 217–228.

Tsui, A. B. M. (1987). An analysis of different types of interaction in ESL classroom discourse. *IRAL*, *25*(4), 336–353.

Uther, M., Zipitria, I., Uther, J., & Singh, P. (2005a). Mobile Adaptive CALL (MAC): A case-study in developing a mobile learning application for speech/audio language training. WMTE'05, pp. 187–191.

Uther, M., Singh, P., Zipitria, I., & Uther, J. (2005b). MAC: An adaptive, perception-based speech remediation s/w for mobile devices. AIED'05.

Uther, M., Singh, P., & Uther, J. (2005c). Mobile Adaptive Call (MAC): A software for speech remediation, In *IEEE Int. Conf. on pervasive services*, 413–416.

Uther, M., Uther, J., Athanasopoulos, P., Singh, P., & Akahane-Yamada, R. (2007). Mobile Adaptive CALL (MAC). A lightweight speech-based intervention for mobile language learners. In *Proc. interspeech 2007 conference*, Antwerp, Belgium, August 27–31, 2007.

Van Patten, B., & Williams, J. (2007). *Theories in second language acquisition: An introduction*. London & Mahwah, New Jersey: Lawrence Erlbaum Associates.

Van Patten, B. (2007). Input Processing in adult second language acquisition. In B. van Patten and J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 115–136). Lawrence Erlbaum.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press.

White, L. (2007). Linguistic theory, Universal Grammar and second language acquisition. In B. van Patten and J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 37–55). Lawrence Erlbaum.

Yanagida, Y., Kawato, S., Noma, H., Tomono, A., & Tesutani, N. (2004). Projection based olfactory display with nose tracking. In *Proc. IEEE virtual reality* (pp. 43–50).

Yang, J. C., Lai, C. H., & Chu, Y. M. (2005). Integrating speech technologies into a one-on-one digital English classroom. WMTE'05, pp. 159–163.