Computer Assisted Language Instruction Consortium

CALICO (ONLINE) ISSN 2056–9017

Article

Learning L2 pronunciation with a mobile speech recognizer: French /y/

Denis Liakin, Walcir Cardoso and Natallia Liakina

Abstract

This study investigates the acquisition of the L2 French vowel /y/ in a mobile-assisted learning environment, via the use of automatic speech recognition (ASR). Particularly, it addresses the question of whether ASR-based pronunciation instruction using a mobile device can improve the production and perception of French /y/. Forty-two elementary French students participated in an experimental study in which they were assigned to one of three groups: (1) the ASR Group, which used an ASR application on their mobile devices to complete weekly pronunciation activities, with immediate written visual (textual) feedback provided by the software and no human interaction; (2) the Non-ASR Group, which completed the same weekly pronunciation activities in individual weekly sessions but with a teacher who provided immediate oral feedback using recasts and repetitions; and finally, (3) the Control Group, which participated in weekly individual meetings 'to practice their conversation skills' with a teacher who provided no pronunciation feedback. The study followed a pretest/posttest design. According to the results of the dependent samples t-tests, only the ASR group improved significantly from pretest to posttest ($p < 0.001$), *and none of the groups improved in perception. The overall success of the ASR group on the production measures suggests that this type of learning environment is propitious for the development of segmental features such as /y/ in L2 French.*

KEYWORDS: LEARNER AUTONOMY; PRONUNCIATION; SECOND LANGUAGE ACQUISITION; SPEECH RECOGNITION

Affiliation

Concordia University, Montreal, Canada.
email: denis.liakin@concordia.ca

CALICO VOL 32.1 2015 1–25

doi : 10.1558/cj.v32i1.25962

Introduction

In the context of mobile devices such as smartphones, tablets, and media players, ASR (Automatic Speech Recognition) is found in the form of applications (apps) which identify the words that a person speaks into a microphone, and automatically convert them into readable text. Recent developments in voice-to-text abilities have encouraged ASR's implementation in computer-assisted language learning (CALL – e.g., Aist, 1999; Cucchiarini *et al.*, 2009; Eskenazi, 1999; Hincks, 2003; Kim, 2006; Neri *et al.*, 2008; Strik *et al.*, 2009). In the context of pronunciation teaching, researchers suggest two possible applications for ASR (Dalby and Kewley-Port, 1999; Holland, 1999; Mostow and Aist, 1999): (1) to teach pronunciation of a foreign language; and (2) to assess students' oral production. These applications have been adopted in a variety of studies on the use of ASR in computer-assisted pronunciation teaching (CAPT) at the segmental level in a second or foreign language (L2) (e.g., Bondar *et al.*, 2011; Cucchiarini *et al.*, 2009; Dalby and Kewley-Port, 1999; Kawai and Hirose, 2000; Kim, 2006; LaRocca *et al.*, 1999; Levis, 2007; Mostow and Aist, 1999; Neri *et al*, 2006, 2008; Penning de Vries *et al.*, 2014; Strik *et al.*, 2009, 2012). Unfortunately, possibly because prosodic information is filtered out in ASR processing, the use of the technology has not received the same level of attention in the investigation of suprasegmental features (Coniam, 2002; Hönig *et al.*, 2012; Kaltenboeck, 2002; Levis, 2007).

One of the interesting aspects of ASR is that it fulfills the criteria proposed by Chapelle and Jamieson (2008) for selecting pronunciation software and activities to develop oral skills. Specifically, ASR allows for: (1) learner fit (ASR is useful for learners as it allows them to identify needed features); (2) explicit teaching (focus on particular pronunciation features and how they contrast with other sounds); (3) opportunities for interactions with the computer, including the ability for learners to speak and analyze their own production; (4) comprehensible and accurate feedback (e.g., visual feedback that uses forms and symbols with which learners are familiar); and (5) the development of strategies for learners to gain an understanding of new features on their own, outside of the language learning or classroom environment.

The main goal of this study is to explore the use of mobile ASR as a pedagogical tool to improve the pronunciation teaching and learning of L2 French. In our investigation, we focus on the acquisition of the French phoneme /y/ (orthographically represented as "u", as in "tu" /ty/ 'you – 2nd person singular') for two main reasons: (1) the sound is very difficult to acquire in both production and perception (e.g., Baker and Smith, 2010; Levy and Law II, 2010; Rochet, 1995); and (2) it has a high functional load (as defined by Brown, 1991 and King, 1967) since it is used to distinguish many highly frequent minimal pairs in French, such as "tu" /ty/ 'you (2nd person singular pronoun)' and

"tout" /tu/ 'all, everything', and "au-dessous" /od.su/ 'below' and "au-dessus" /od.sy/ 'above'.

To our knowledge, there are no studies that have investigated the use of ASR on mobile devices for pronunciation teaching and/or learning (see also Godwin-Jones (2009) for a similar observation), including the development of production and perception.

Background

ASR and the Acquisition of Second Language Pronunciation

There are three categories of ASR systems (Rosen and Yampolsky, 2000; Young and Mihailidis, 2010), which are differentiated by the degree of user training required prior to use: (1) speaker dependent; (2) speaker independent; and (3) speaker adaptable. Speaker dependent ASR requires the user to train the speech recognizer with samples of his/her own speech; consequently, the system works well only for the person who trains it. Speaker independent ASR does not require speaker training prior to use because the recognizer is pretrained during system development with speech samples from a variety of speakers. Many different speakers will thus be able to use the same ASR application with relatively good accuracy as long as their speech falls within the range of the collected samples. Speaker adaptable ASR is similar to speaker independent ASR in that no initial speaker training is required prior to use. However, unlike speaker independent ASR systems, as the speaker adaptable ASR system is used over time, the recognizer gradually adapts to the speech of the user.

Another way of characterizing ASR technology is by the type of input that the system can handle: (1) isolated/discrete word recognition; (2) connected word recognition; and (3) continuous speech recognition (see Jurafsky and Martin, 2008; Rabiner and Juang, 1993; Rosen and Yampolsky, 2000). Discrete word recognition requires a pause or period of silence to be inserted between words or utterances. Connected word recognition is an extension of discrete word recognition and requires a pause or period of silence after a group of connected words have been spoken. In continuous speech recognition, an entire phrase or complete sentences can be spoken without the need to insert pauses between words or after sentences. Different combinations of these ASR types have been utilized in the CAPT/ASR literature. In our experiment, we adopted a speaker independent system designed for continuous speech recognition, as will be described in the forthcoming methodology section.

The majority of the studies that have investigated the effects of ASR on the acquisition of L2 pronunciation have shown that, despite many limitations, this technology has the potential to be effective. Early explorative work by Dalby and Kewley-Port (1999), LaRocca *et al.* (1999), and Mostow and Aist

(1999) indicated that ASR technology was still not as accurate as human analysis and, consequently, they suggested that the software could be useful for student practice with only certain aspects of pronunciation: segmental features. Recent developments in ASR designed particularly for language learning have shown the effectiveness of the technology for L2 pronunciation training (e.g., Cucchiarini *et al*., 2009; Neri *et al*., 2008; Strik *et al*., 2009).

A different type of ASR can be found in the form of dictation software, which are computer applications that allow users to speak freely as the application transcribes what they say (e.g., NCH Express Dictate, Nuance Dragon Speech Recognition). Although not as commonly researched, early studies using this off-the-shelf technology include those of Coniam (1999), and Derwing *et al.* (2000), who evaluated its recognition performance for English. The authors demonstrated that while dictation software offered positive results for native English speakers (90% accuracy), it performed less well for non-native speakers, leading the authors to conclude that the technology was not mature for use in L2 learning.

A critique of this type of ASR appeared in Neri *et al.* (2003), where the authors described the inadequacies of dictation software: that the technology was developed to recognize native speech only and, as such, did not include any mechanism to provide feedback on pronunciation quality. Accordingly, these dictation packages performed poorly with non-native speakers because of the acoustic variations found in their speech. Specially-designed ASR systems, on the other hand, have better recognition performance with non-native speech because their underlying acoustic models are prepared to accept the mispronunciations that language learners are expected to make.

In sum, the available literature suggests that ASR technology may have positive effects on the acquisition of L2 pronunciation. With regards to dictation software, despite the pessimistic results obtained in the studies conducted over a decade ago, our experience with current ASR systems suggests that the technology has advanced considerably in the detection of non-native speech. We thus hypothesize that ASR software designed for dictation could be beneficial for pronunciation training, and that learners will also benefit from the technology if it is offered in a portable format.

Mobile technology and second language acquisition

The use of mobile devices such as smartphones, media players and camcorders for language learning has sparked the interest of an increasing number of researchers over the last decade, particularly in the field of *vocabulary* acquisition (e.g., Kiernan and Aizawa, 2004; Kennedy and Levy, 2008; Lu, 2008; Zhang *et al.*, 2011). Despite being considered by teachers and parents as a distraction in the classroom, these studies suggest that mobile devices can be

useful for language learning. In addition, their multimedia capabilities can help students have more authentic learning experiences, situating learning within their cultural and linguistic schemata (Joseph and Uther, 2009).

Despite encouraging results, Kukulska-Hulme and Shield (2008) observed that Mobile-Assisted Language Learning has not yet been embraced on a large scale and has not yet received sufficient research attention toward its full potential as a pedagogic practice. Along the same lines, Joseph and Uther (2009) stressed that the value of using mobile devices and incorporating multimedia elements into language learning applications needs to be quantified with controlled experiments, where the control groups study on non-mobile platforms or in mobile contexts with non-technical support, e.g., via paper flashcards. According to these two authors, experiments of this sort should be prioritized in future research. The current study addresses this recommendation by incorporating a control and a comparison (teacher-driven) group with characteristics similar to what these authors recommended.

Consistent with Godwin-Jones' (2009) observation, as indicated earlier, we are not aware of any study that investigates the use of ASR on mobile devices and its effects on L2 pronunciation. To assess the viability of using mobile ASR technology and to test its effects on learning, we chose to focus on the acquisition of L2 French /y/.

French /y/ and its acquisition: Production, perception, and functional load

The target French pronunciation feature examined in this study was the vowel /y/. This is an ideal target phoneme for pronunciation instruction because, as mentioned earlier, /y/ is highly problematic for L2 learners in both production and perception (Baker and Smith, 2010; Levy and Strange, 2008). One possible explanation for why this phoneme is so difficult to acquire might be due to its perception by L2 learners whose languages lack /y/ in their phonological inventories.

In first language acquisition, the most accepted assumption is that perception must precede production, because while children are assumed to quickly develop adult-like competence (Hale and Reiss, 1998; Stampe, 1973), their articulators do not follow the same rate of development. For L2 acquisition, on the other hand, the issue is not as straightforward, and has led researchers to stand on one or more sides of three logical hypotheses regarding the relationship between perception and production: (1) perception precedes production (e.g., Flege, 1995; Borden *et al.*, 1983); (2) production precedes perception (e.g, Sheldon and Strange, 1982; Sheldon, 1985); and (3) production and perception develop simultaneously (e.g., Flege, 1999; Koerich, 2006).

The common denominator among these hypotheses is that they recognize the importance of L1 phonotactic knowledge, against which all foreign features are 'filtered' and subsequently categorized into a language system (interlanguage). This notion has given rise to at least two models for speech perception and L2 learning in general: Flege's (1995) Speech Learning Model (SLM) and Best's (1993; 1995) Perceptual Assimilation Model (PAM). Leaving aside conceptual differences between these models, Flege's SLM model postulates that phonetically similar L2 sounds are more likely to be perceived via the L1 than those that are dissimilar, possibly due to perceptual salience in the latter case. In the case of similar sounds, the foreign segment (or phonetic feature) is subsumed within the existing perceptual representation for a comparable sound in the L1, which as a result leads to a so-called foreign accent. Similar predictions are also made by Best's PAM model, which proposes that 'non-native segments […] tend to be perceived according to their similarities to, and discrepancies from, the native segmental constellations that are in close proximity from them in phonological space' (p. 193). In Best's view, novel sounds are assumed to be either assimilated to a native L1 category (in the case of similar sounds) or to an uncategorizable sound that will form a new category (in the case of dissimilar sounds). To summarize, these two models predict that new L2 segments that are perceptually similar (assimilable) will be of greater difficulty to acquire than those that are dissimilar (unassimilable).

It is premature and beyond the scope of this investigation to provide a definite answer to the question regarding the nature of French /y/ as 'filtered' by the different L1 phonologies included in this study (see the method section): Does the representation of French /y/ pattern with the *similar* or the *dissimilar* scenarios predicted by the SLM and PAM models? However, based on the high degree of difficulty that many French L2 learners have in acquiring /y/ (e.g., Baker and Smith, 2010; Levy and Law II, 2010; Rochet, 1995), and on the fact that the L1s considered in this study have equivalent (and 'assimilable') /y/s (e.g., /u/ for English, Farsi, and Spanish speakers, and /i/ for Portuguese speakers – see forthcoming discussion), it is reasonable to conjecture that /y/ can be subsumed under the *similar* pattern. Accordingly, these L1 speakers will categorize French /y/ based on their L1 phonotactic knowledge of a similar L1 phoneme: /i/ or /u/.

In addition to its difficulty in production and perception, French /y/ has a high functional load (King, 1967), a concept used to describe the extent and degree of contrast between linguistic units, usually phonemes. In phonology, it is a measure of the work that two phonemes do to maintain phonemic contrast in all possible environments, involving minimal pairs whose members are both frequent (Brown, 1991). Consequently, certain phonemes in a language have a higher functional load than others depending on the degree to which

they contrast in meaning. For instance, French /u-y/ is used to distinguish highly frequent French minimal pairs such as *au-dessous* /odsu/ 'below' from *au-dessus* /odsy/ 'above', a type of alternation that, due to its high frequency, may affect intelligibility. Because many languages lack this highly functional phoneme, it is essential that it be mastered early on in order to not compromise meaning in the target language. This is one of the arguments that Jenkins (2000; 2002) used in her rationale for proposing her version of the English as a Lingua Franca approach, particularly in deciding priorities for pronunciation teaching. According to the author, priority should be given to sounds that have a high functional load, a requirement which we believe is fulfilled by French /y/.

Research questions and predictions

The purpose of the present study is to investigate the acquisition, in terms of production and perception, of the French vowel /y/ in a mobile ASR-based learning environment. It thus aims to examine the feasibility of using mobile ASR as a pedagogical tool for L2 pronunciation learning. Accordingly, the following two research questions guided our investigation:

1. Does ASR-based pronunciation practice using a mobile device improve French L2 /y/ *production*?
2. Does ASR-based pronunciation practice using a mobile device improve French L2 /y/ *perception*?

On the basis of the research discussed earlier, we hypothesized that ASR would have a positive effect on /y/ production, since an explicit focus on pronunciation in an ASR-based environment may improve learners' production. This assumption is consistent with the works of Neri *et al.* (2008) and Cucchiarini *et al.* (2009), among others. With respect to the second question, we predicted that learners would be able to extend the newly acquired productive skill into perception, as has been attested (but less commonly) in the literature (e.g., Aliaga-Garcia and Mora, 2009; Bradlow *et al.*, 1997; Jongman and Wade, 2007; but note that these studies focus on the effects of phonetic training on the development of perceptual skills). For the sake of this study, we define perception as the participant's ability to discriminate between a set of options, namely /y/, /u/ and /i/, embedded in words, phrases and sentences, as will be discussed later.

Method

Participants

Forty-two adult students of French as a second language participated in this study, with an average age of 22 (30 female, 12 male). All participants were recruited from three intact L2 French classrooms at two Anglophone universities in Montreal. They were either native English speakers or had native-like

proficiency in the language (English: *n* = 27, Farsi: *n* = 2, Spanish: *n* = 7, Brazilian Portuguese: *n* = 2, Chinese: *n* = 2, Serbian: *n* = 1, Japanese: *n* = 1). All participants had elementary-level proficiency in French (A2 level, according to the Common European Framework of Reference for Languages – this is a requirement for enrolment in the 'Elementary French' classes from which the participants were recruited) and, accordingly, had not yet fully acquired the target phoneme /y/. Because of these requirements (i.e., A2 level proficiency and performance on the pretest), data from eight students were discarded because they scored more than 50% accuracy in /y/ production and perception in the pre-test.

Design of the study, experimental groups, and treatment

Following Chapelle's (2001, 2012) recommendation for conducting methodologically convincing CALL research, this study followed a mixed-methods approach, using a pre/post research design (quantitative) followed by surveys and interviews with the participants (qualitative). Due to the scope of this study and its main goals, the focus will be primarily on the analysis of the quantitative results.

The 42 participants recruited for this study were randomly assigned to one of three distinct groups, each corresponding to an experimental group: ASR, NASR (Non-ASR) and CTL (Control). During the treatment period, the participants were not informed about the nature of study, except that it was about 'an app that could help second language learners improve their French'. Figure 1 illustrates the general design of this study, which will be discussed in detail below.

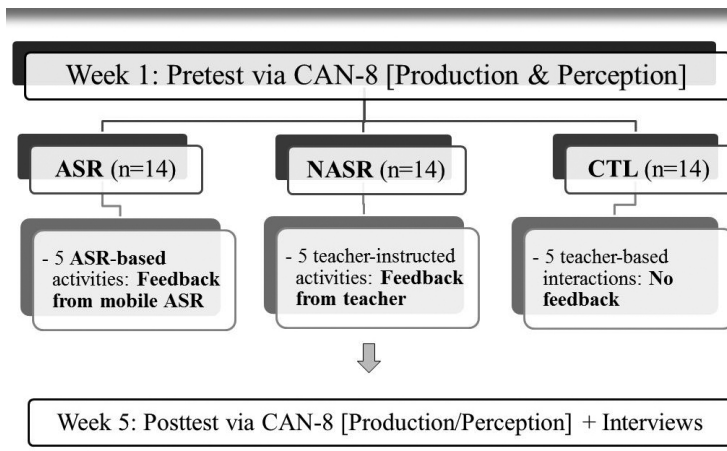

Figure 1. Design of the study. ASR = Automatic Speech Recognition group; NASR = Non-ASR group; CTL = Control group.

The ASR Group corresponded to the group that practiced French pronunciation using mobile ASR on an iPod Touch or iPhone using a commercial (but free) ASR application: Nuance Dragon Dictation, a speaker independent dictation system designed for continuous speech recognition, as described earlier. To test the accuracy of the application, five native French speakers pronounced all words/phrases included in the study using the university's wi-fi connection. One hundred percent of utterances were recognized correctly. The students completed on a weekly basis, either at home or at the university, five 20-minute pronunciation activities that consisted of reading aloud the target words and phrases in French using the ASR software installed on their mobile devices. After each reading attempt, students were provided with immediate written visual feedback via an orthographic representation of their attempt. To illustrate, if students attempted to pronounce the word 'pure' [pyr] and 'pour' or 'pire' appeared on their screen as the written (visual) result, this indicated that their pronunciation was incorrect, thus requiring another attempt. In some cases, a slow internet connection or background noise would affect the results, but students were aware of this limitation and were therefore asked to be patient, try again on another network, or wait until they were on university premises for a faster and more reliable wireless connection. The ASR participants were asked to spend approximately one minute per word/phrase, depending on the level of difficulty of each target phrase, for a total of 20 minutes. To ensure that the participants completed the assigned ASR-based pronunciation activities, they were also asked to indicate, on a 'pronunciation form' (see Appendix), the number of times they repeated each form until they were able to produce it accurately or until their one-minute limit had expired.

The 'Non-ASR Group', on the other hand, did not have access to mobile ASR. However, they completed the same activities that the ASR participants did: They read aloud the same words and phrases in individual, weekly 20-minute sessions with a French teacher who provided immediate oral feedback on their pronunciation using recast and repetitions. To accommodate the nature of the intervention received by this group, the pronunciation form (Appendix) was slightly adapted (e.g., the irrelevant question 'what word/s do you see on the screen?' was removed). For comparable treatments, the teacher was asked not to volunteer any pronunciation practice that emphasized the target phoneme /y/; instead, the teacher was asked to concentrate on the items listed on the form and provide only what feedback was necessary for the completion of the activities.

Finally, the 'Control Group' participated in weekly individual 20-minute meetings with the goal of practicing their conversation skills with a French teacher who provided no feedback on /y/ pronunciation. These sections could be described as conversation classes, in which the participant and the teacher engaged in discussions of a variety of topics about school, aspirations, family, etc.

Table 1 illustrates the activities accomplished throughout the duration of the study, the focus of instruction, the type of feedback provided in each group, and the length of each corresponding treatment.

Table 1. Experimental groups and related activities

	Experimental groups		
	ASR	NASR	Control
Activity	Oral ASR activities	Oral teacher-based activities	Conversation with teacher
Focus	/y/ + distractors	/y/ + distractors	None
Feedback	Mobile ASR (written)	Teacher (oral)	None
Length	Five 20-min weekly sessions	Five 20-min weekly sessions	Five 20-min weekly sessions

Tasks: Pretest and posttest

For the production and perception tasks used to measure students' pronunciation capabilities, we employed CAN-8 VirtuaLab, 'an interactive, multimedia tool used for the instruction of modern languages' with which the participants were familiar (they used it on a weekly basis in the university's language lab to complete general language activities).

The production task consisted of reading words and phrases aloud, which were recorded automatically using CAN-8 in the university's language lab, without the presence of the researcher or teacher. We targeted 20 instances of /y/ (plus 15 distractors) in 19 keywords (one containing two instances of the target phoneme) which were carefully selected so that /y/ occurred in equally distributed syllabic environments: 10 in open, vowel-final syllable structures (e.g., -du [dy] in 'defendu'), and 10 in closed, consonant-final syllable contexts (e.g., cul- [kyl] in 'culture'). The words selected for the production task were: *assume*, *azur*, *chute*, *culture*, *défendu*, *fumes*, *lune*, *musique*, *numéro*, *particule*, *perdu*, *plu*, *pulvérise*, *surtout*, *tu*, *ultime*, *unanime*, *une*, and *vu*. Figure 2 shows the CAN-8 interface illustrating a production task:

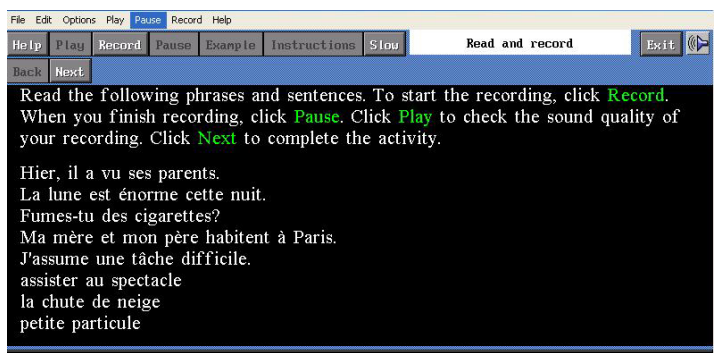

Figure 2. The production task: An example.

In the perception task, we employed pseudowords to avoid frequency and familiarity effects; this was based on the assumption that the productivity of a pattern is often determined by its frequency in the input, i.e., 'the more items encompassed by a schema, the stronger it is, and the more available it is for application to new items' (Bybee, 2001: 13; see also Flege *et al.*, 1996 for similar assumptions). To illustrate a possible frequency (and consequently a familiarity) effect in the context of the study, some participants could select the French word 'tu' [ty] as containing /y/ simply due to their familiarity with the word as a consequence of its high frequency in their language input (and possibly in their language output).

In the perception experiment, the participants listened to 45 monosyllabic 'French' pseudowords containing the vowels /y/, /u/ and /i/ (15 instances of each vowel; e.g., fuppe [fyp], foupe [fup], fippe [fip]). The vowels /u/ and /i/ were included as distractors even though they have been reported as the most confusable vowels in the identification of French /y/. For instance, while Anglophone listeners perceive the French /y/ as their L1 back /u/ vowel (Gottfried, 1984; Rochet, 1995), Brazilian Portuguese and Haitian Creole listeners perceive the same vowel as their own /i/ (Rochet, 1995).[1]

The perception task followed a four-item multiple-choice format, with each alternative representing one of the relevant three vowels described above, and 'I don't know' as the fourth choice to minimize random selection. After listening to a pseudoword such as *fuppe* [fyp], participants were asked to choose the alternative that corresponded to what they heard. Figure 3 illustrates the interface of the perception task on CAN-8.

Figure 3. Perception task: An example.

Analysis

To assess the students' production, two bilingual francophone RAs listened to each student's recordings and determined whether the pronunciation of /y/ was correct or incorrect. In the case of conflicting opinions, a member of our team listened to those occurrences and made the decision. If an item was ambiguous or unclear, that item was excluded from the computation. In total, there were 1,680 occurrences of /y/ with an inter-rater reliability of 88.7% (1,490/1,680). Assessment of the students' perception was done automatically by CAN-8, which is programmed to assess each response as correct or incorrect according to the stimulus input into the system.

For the statistical analysis of the data and to test for differences among the three groups on the pretest and posttest, a one-way ANOVA was performed at each time for production and perception. To test for differences within each group over time, dependent sample *t*-tests were carried out to compare the pretest to posttest performances for each group.

Results

The general descriptive statistics of the analysis for /y/ production and perception appear in Table 2. The mean scores (M) of accurate production and perception are presented as well as the standard deviations (SD) across the two tests (pretest and posttest) and the three groups under consideration (ASR, Non-ASR and Control). Because there were ten tests performed, the alpha level had to be adjusted and set at 0.005 (0.05/10 tests). Overall, the results of the one-way ANOVA indicated no differences among the three groups either on the pretest or the posttest in both /y/ production (F (2, 39) = 0.95, p = 0.392 and F (2, 39) = 0.90, p = 0.413 in pre- and posttest respectively) and /y/ perception (F (2, 39) = 1.57, p = 0.221 and F (2, 39) = 0.32, p = 0.731 in pre- and posttest respectively).

Table 2. Descriptive statistics for /y/ production and /y/ perception over time, across the three groups (Mean scores)

	Production (n = 20)						Perception (n = 15)					
	ASR		Non-ASR		Control		ASR		Non-ASR		Control	
Test	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pre	7.09	5.51	9.79	3.98	8.43	5.86	8.07	3.64	9.64	3.67	10.29	2.81
Post	10.71	4.23	11.86	3.98	9.50	5.53	9.93	2.89	10.29	3.77	10.93	3.40

To test for differences within each group over time, dependent samples *t*-tests were carried out comparing pretest performance to posttest performance for each group. In this analysis, only the ASR Group improved

significantly from pretest to posttest in /y/ production ($p < 0.001$), and no group improved in /y/ perception.[2] The following two sections will provide a detailed report of each of these sets of results.

Production of French /y/

The first research question asked whether ASR-based pronunciation practice using a mobile device would improve French L2 /y/ *production*. According to the results of the dependent samples *t*-tests, only the ASR group improved significantly from pretest to posttest ($p < 0.001$). This indicates that learners who received instruction via the mobile ASR application displayed more improvement over time than those who received teacher-based input and feedback (Non-ASR) or no input or feedback whatsoever (Control). As such, these results support our initial hypothesis that the pedagogical use of a mobile version of ASR would have a positive effect on /y/ production.

For illustrative purposes, the results for production are presented in Figure 4, where the mean scores for accurate /y/ production are presented.

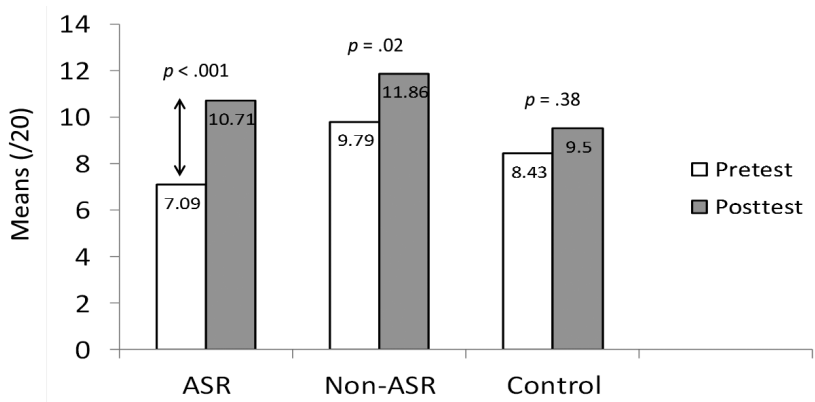

Figure 4. /y/ production results

Perception of French /y/

The second research question asked whether ASR-based pronunciation practice using a mobile device would improve French L2 /y/ *perception*. The results of the dependent samples *t*-tests, illustrated in Figure 5, indicate that despite slightly greater gains for the ASR group, the three groups behaved in a similar way (pre/posttest differences: ASR: $p > 0.05$; Non-ASR: $p > 0.38$; Control: $p > 0.37$). This indicates that the group that received ASR-based treatment was not able to extend the newly acquired knowledge detected in production to perception; accordingly, our initial hypothesis was not supported.

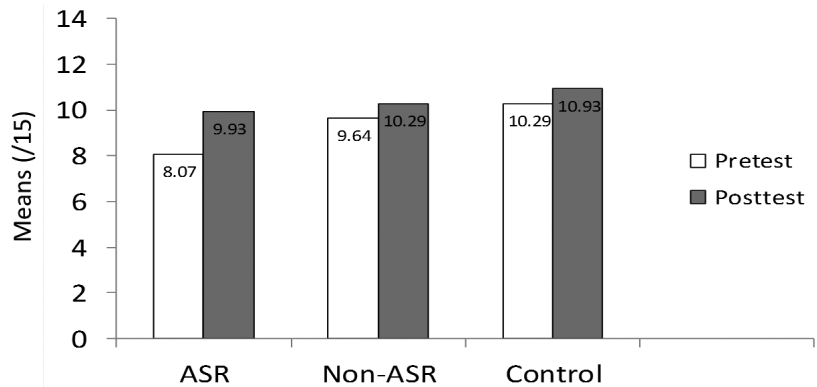

Figure 5. /y/ perception results

Discussion

The main goal of this study was to explore the use of ASR software on mobile devices as a pedagogical tool for improving L2 French pronunciation in production and perception. With regards to production, the results indicate that, similar to what is observed in the general (non-mobile) ASR literature, the use of mobile speech recognition appears to have a positive effect on the acquisition of the French vowel /y/ (see also Cucchiarini *et al.*, 2009 and Neri *et al.*, 2008 for similar results involving segments). We attribute these learning gains to a variety of factors that include insights from the general SLA/CALL literature, notably Chapelle's (2001) ideas about input enhancement and computer-aided interaction (e.g., /y/ pronunciation was reinforced via orthography, input manipulation and repetition among ASR users), the effects of an explicit focus on the target form (Dabaghi, 2010; Dekeyser, 1993), immediate feedback (Rosa and Leow, 2004), multiple opportunities for learning (Christison, 1999; Chun and Plass, 1996), and the game-like approach to teaching afforded by mobile technologies (Bruff, 2009). Lastly, mobile ASR technology, as utilized in this study, ascribes to Chapelle and Jamieson's (2008) suggestions for selecting pronunciation software to develop speaking skills, based on research by Hardison (2004; 2005), Derwing *et al.* (1998), and MacDonald *et al.* (1994). Accordingly, the mobile ASR technology adopted in this study provides: learner fit (it emphasizes a feature that the participants needed to improve); potential for explicit teaching and learning; opportunities for interactions with the computer; comprehensible and visual (orthographic) feedback; and strategy development to guide students to start learning new L2 features on their own outside of the language learning environment. Evidently, we are aware that the observed gains could also be caused by the effect of the adoption of a new technology, which may have increased the overall interest and motivation of the students (Clark, 1983; Strambi, 2001; Warchauer, 1996).

Regarding perception, our results indicate that L2 learners were not able to transfer the acquired knowledge about /y/ production into perception. We attribute this result to at least two main factors. First, it is possible that the total of 1.5 hours of instruction were not sufficient for learners to acquire /y/ in perception and thus locate this foreign phoneme within the phonological system that characterizes their L1s. As discussed earlier, this phoneme is highly complex from both an acoustic and articulatory perspective (Baker and Smith, 2010; Levy and Strange, 2008); this may affect its acquisition in perception, particularly in an experiment in which no emphasis was given to the development of perceptual skills. Secondly, we admit that we were originally optimistic about our conjecture that a focus on production could translate into gains in perception, as has been argued in studies that focus on the effects of phonetic training on the development of perceptual skills (e.g., Aliaga-Garcia and Mora, 2009; Bradlow *et al.*, 1997). Instead, the results obtained in our study seem to conform to those related to the acquisition of /r/ and /l/ by Japanese learners of English (e.g., Hattori, 2009; Sheldon and Strange, 1982). In these studies, L2 learners were able to produce these two English liquids more reliably than they were able to perceive them. In Hattori (2009), for instance, Japanese learners could be trained to produce native-like English /r/ and /l/ approaching 100% accuracy, while the same learners could not distinguish between the two phonemes in perception experiments. In sum, along the lines of Hattori (2009) and Sheldon and Strange (1982), our findings seem to suggest that speech production can sometimes precede its perception (at least for the acquisition of /y/ in a ASR-based context), as the participants in the ASR group improved only in the former. However, based on the general trends observed (e.g., the ASR group did outperform the other two groups, but not significantly), the results point in an optimistic direction regarding the use of mobile ASR for the development of speech perception.

Concluding remarks

The present study revealed a significant improvement in /y/ production by the group that trained in an ASR-based environment. The overall success of this group on the production measures suggests that this type of learning environment is propitious for the development of L2 French /y/ and, we speculate, for the development of other related segmental features. This has both theoretical and practical relevance.

With regard to its theoretical contribution, albeit limited in scope, the study initiates a debate on the potential and feasibility of using mobile ASR technology for the teaching and learning of L2 segments, particularly French /y/. In addition, the results obtained reinforce some of the well-established notions instituted in the CALL and SLA literature in the context of mobile technology.

As discussed earlier, these include learner autonomy (Holec, 1981; Schwienhorst, 2008), immediate feedback (Rosa and Leow, 2004), explicit instruction (Dekeyser, 1993), input enhancement (Chapelle, 2001), and multimodal exposure to the forms being acquired (Christison, 1999).

From a pedagogical standpoint, we believe that ASR software on mobile technology should be further explored as a potential *complement* to pronunciation activities conducted in language classrooms: It may not only promote the acquisition of segments, as demonstrated in this study, but it can also be used by teachers and students without much preparatory work (contrary to the types of specially-designed ASR systems used in studies such as those of Neri *et al*., 2006), and it provides a type of feedback that is easily understood, via orthography. In the classroom, teachers could emphasize meaningful communicative tasks, as recommended by L2 pedagogues (e.g., Littlewood, 2004; Nunan, 2004), while assigning certain pronunciation tasks (for instance, those that are repetitive and require a special focus on articulation) as personalized homework assignments. Those tasks could target particular pronunciation problems such as French /y/: it is difficult to produce and perceive; it requires the articulation of 'funny lip-rounding' which may inhibit shy students in public environments; and it has a high functional load, meaning that its mispronunciation has high chances of affecting intelligibility. Accordingly, we believe that ASR can and should be used in the language learning environment because: (1) it has the potential to improve L2 learners' pronunciation, as we have shown here; (2) it can relocate resources so that classroom time can be used exclusively (or mostly) for communicative activities; (3) it accommodates a wide variety of learners (e.g., those who benefit from the visual interactions afforded by ASR; Gardner, 1983); and, finally, (4) the technology was evaluated very positively by the participants, as indicated by the following samples of participants' responses:

- It is perceived as having a positive effect on pronunciation ('[…] *will help you pronounce better*'; '*They should definitely implement that in the grammar classes because it's like you need to know how to pronounce things*');
- It provides immediate visual feedback ('*You pronounce and you see right away what you're pronouncing*'; '*It gives you the answers, you can see*');
- It is portable and convenient ('*It's good to have homework that you can take home and practice pronouncing on your own, instead of just in the lab*');
- It provides a different modality to learn ('*So, yeah. Especially when there's no one else, [these are] different exercises. It's better, yeah*'); and

It encourages practice ('*I get nervous when it's, like, in person. So, it's definitely easier, and then I can get more comfortable, and I don't mess up so much, in person if I … you just get more confident*').[3]

We are aware that it is premature to arrive with certainty at generalizable conclusions in a study of such narrow scope (e.g., focus on a single phoneme, involving participants recruited from three intact classes in two universities) and in a field that is still in its infancy (mobile ASR-based technology for pedagogical purposes). As such, there are some limitations that will deserve special consideration in future investigations. One of the major limitations of this study, as alluded to above, is its limited contribution to the field, as it investigates the acquisition of a single phoneme in French: /y/. Two important questions remain for further investigation: Will other phonological or phonetic items such as features (e.g., spread glottis, voice-onset time), syllable structure (e.g., codas), rhythm and intonation benefit from a similar (mobile) ASR treatment? What is the impact of ASR-based training on overall pronunciation skills (e.g., the development of intelligibility)? Another limitation of this study relates to what is referred to as the novelty effect. As has been attested in the computer-assisted learning literature (e.g., Nikolova, 2002; Warschauer, 1996), there is the possibility that the gains observed in the ASR group are ephemeral, merely a reflection of what Clark (1983) defines as the novelty effect, wherein it is assumed that the improved performance observed is a response to the increased interest in the new technology, and not necessarily a direct influence of its use. Similarly, it is also possible that the improved learning observed in the ASR Group was affected by the instructional methods associated with the use of this new technology, as discussed above (e.g., the development of learner autonomy, encouragement of repetition, presence of immediate feedback and multimodal exposure). Only an extensive longitudinal study, conducted after the novelty factor has worn off, will be able to address this concern.

Some of the methodological limitations of this study include: the short duration of the treatment and training sessions, the linguistic heterogeneity of the three groups of participants whose first languages differed (while most were native English speakers, some were multilingual), and the small number of participants. This latter limitation was mostly due to the fact that the majority of our participants did not own an appropriate device to participate in the study and, to a lesser extent, to participant attrition (one participant withdrew from the experiment due to illness).

A potential direction for future research is to adapt and/or develop mobile technologies that can address the full spectrum of what linguistic competence in an L2 entails. According to Dickerson (2004, 2013), this knowledge includes learners' ability to perceive (e.g., distinguish /u/ from /y/), produce

(e.g., articulate /y/), and predict pronunciation patterns (based on grapheme-to-phoneme rules: e.g., while orthographic 'u' is pronounced as [y], 'ou' and its orthographic variants such as 'oup' and 'out' are produced as [u]). These three competence elements or 'trilogy of goals' (prediction, production, perception), can be easily explored in a mobile-assisted learning environment via a combination of tools/apps that promote the development of production (ASR), perception (text-to-speech synthesizers – TTS),[4] and prediction (ASR and TTS).

Notes

1. These types of patterns and observations led many L2 researchers to propose different models for second language speech perception. The two most notable ones are Flege's (1995) Speech Learning Model, discussed earlier, and Best's (1993, 1995) Perceptual Assimilation Model. For Best (1995), 'non-native segments … tend to be perceived according to their similarities to, and discrepancies from, the native segmental constellations that are in close proximity from them in phonological space' (p. 193).

2. A finding of no significant difference on the posttest does not mean that there could not be a significant change over time in each group. In our case, the interpretation of the results is a bit more subtle. The ASR group had the lowest mean at the pretest, but with such large within-group variability, as evidenced by the standard deviation, no significant difference was found between it and the other two groups. On the posttest, the ASR group did show the greatest gain in mean score and its within-group variability had decreased somewhat too. One could argue that this result was caused by the fact that the ASR group had more room to improve, since it had the lowest (but not significantly different) mean at the pretest. We must note, however, that the pretest to posttest movement for the other two groups was not limited by a ceiling effect. Only one subject in these two groups achieved 19 out of 20 on the posttest. The next highest score in both of these groups was 17 out of 20. Examining individual scores, the more dramatic change for the ASR group may be attributed to the huge change in scores for four of five very low scorers on the pretest (1 > 10, 2 > 10, 1 > 9, 2 > 7). Only one low scorer in the other two group made similar improvement from pretest to posttest (Non-ASR: 1 > 9; Control: 4 > 9).

3. According to the participants, the two main weakenesses of the ASR system adopted are the unreliability of the internet connection, particularly during peak periods and in weak spots within the university premises (e.g., "*Sometimes the app wouldn't work because of my bad connection and I'd get frustrated*"), and the level of pronunciation accuracy required by the app (e.g., "*Sometimes I would pronounce it correctly, and my boyfriend is from Paris, and he would say "yeah, that's right", and then he would say it and it still didn't come out*."). We suspect that comments similar to the latter can sometimes be attributed to the effects of a faulty internet connection.

4. A text-to-speech synthesizer (TTS) is a computer program/app that generates speech from any written text automatically. TTS programs feature different speed levels of the speech output, both female and male speakers with different pitches (low and high), different accents of language varieties, and a highlight function that displays the words, sentences and paragraphs being read by the program in color. The quality of the synthesis has improved substantially over the years (Handley, 2009), and we believe that this is an appropriate time to start exploring this computer application, in a mobile environment, as a potential model for L2 speech. The main advantage of TTS is that it can be used as a means of enhancing the L2 aural input and, therefore, help learners perceive some of the phonetic properties of /y/ as well as the acoustic differences between this vowel and equivalent L1 forms such as /u/ and /i/. This could potentially help learners decipher the intricacies of assigning a 'phonological space' to foreign /y/ in their developing phonologies.

About the authors

Denis Liakin is an Associate Professor of French and Linguistics at Concordia University (Montreal, Canada). His research interests include effects of computer technology on L2 learning, corrective phonetics and second language acquisition of syntax.

Walcir Cardoso is an Associate Professor of Applied Linguistics at Concordia University (Montreal, Canada). He conducts research on the second/foreign language acquisition of phonology, morphosyntax and vocabulary, and the effects of computer technology (e.g., clickers, text-to-speech synthesizers, automatic speech recognition) on L2 learning.

Natallia Liakina's professional experience includes teaching French as a second language at the university level in Ontario and in Quebec. Since 2006, she has taught at the French Language Centre at McGill University. Her current research is focused on corrective phonetics and the impact of new technologies on second language teaching and learning both in the classroom and in computer lab settings.

References

Aist, G. (1999). Speech recognition in computer-assisted language learning. In K. Cameron (ed.), *CALL: Media, Design & Applications,* 165–181. Lisse, Holland: Swets & Zeitlinger.

Aliaga-Garcia, C. and Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In B. Baptista, A. Rauber and M. Watkins (eds), *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, 2–31. Newcastle Upon Tyne: Cambridge Scholars.

Baker, W. and Smith, L. (2010). The impact of L2 dialect on learning French vowels: Native English speakers learning Québécois and European French. Canadian. *Modern Language Review*, 66 (7): 711–738. http://dx.doi.org/10.3138/cmlr.66.5.711

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schoenen, P. Jusczyk, P. MacNeilage and J. Morton (eds), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, 289–304. Dordrecht: Kluwer Academic Publishers.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In: W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171–206. Baltimore, MD: York Press.

Borden, G., Gerber, A. and Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning* 33 (3): 499–526. http://dx.doi.org/10.1111/j.1467-1770.1983.tb00946.x

Bradlow, A. R., Pisoni, D. B., Yamada, R. A. and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: II. Some effects of perceptual learning on speech

production. *Journal of the Acoustical Society of America*, 101 (4): 2299–2310. http://dx.doi.org/10.1121/1.418276

Brown, A. (1991). Functional load and the teaching of pronunciation. In A. Brown (ed.), *Teaching English Pronunciation: A Book of Readings,* 211–224. London: Routledge.

Bruff, D. (2009). *Teaching with Classroom Response Systems: Creating Active Learning Environments*. San Francisco, CA: Jossey-Bass.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511612886

Chapelle, C. (2001). *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing, and Research.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524681

Chapelle, C. (2012, April). Using mixed-methods research in technology-based innovation for language learning. Paper presented at the Innovative Practices in Computer Assisted Language Learning Conference, University of Ottawa, Ontario.

Chapelle, C. and Jamieson, J. (2008). *Tips for Teachers: Computer-assisted Language Learning*. New York: Pearson Longman.

Chun, D. M. and Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80 (2): 183–198. http://dx.doi.org/10.1111/j.1540-4781.1996.tb01159.x

Christison, M. A. (1999). *A Guidebook for Applying Multiple Intelligences Theory in the ESL/EFL Classroom*. Burlingame, CA: Alta Book Center Publishers.

Clark, R. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53 (4): 445–459. http://dx.doi.org/10.3102/00346543053004445

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English, *System* 27 (1): 49–64. http://dx.doi.org/10.1016/S0346-251X(98)00049-9

Cucchiarini, C., Neri, A. and Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, 51 (10): 853–863. http://dx.doi.org/10.1016/j.specom.2009.03.003

Dabaghi, A. (2010). *Corrective Feedback in Second Language Acquisition: Theory, Research and Practice*. LAP Lambert Academic Publishing.

Dalby. J. and Kewley-Port, D. (1999). Explicit pronunciation training using automatic speech recognition. *CALICO Journal* 16 (3): 425–445.

Dekeyser, R. M. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *The Modern Language Journal,* 77 (4): 501–514. http://dx.doi.org/10.1111/j.1540-4781.1993.tb01999.x

Derwing, T., Munro, M. and Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning,* 48 (3): 393–410. http://dx.doi.org/10.1111/0023-8333.00047

Derwing, T., Munro, M. and Carbonaro, M. (2000). Does popular speech recognition

software work with ESL speech?, *TESOL Quarterly* 34: 592–603. http://dx.doi.org/10.2307/3587748

Dickerson, W. (2004). *Stress in the Speech Stream: The Rhythm of Spoken English*. Urbana, IL: University of Illinois Press.

Dickerson, W. (2013). Prediction in teaching pronunciation. In C. Chapelle (ed.), *The Encyclopedia of Applied Linguistics.* Oxford: Wiley-Blackwell.

Eskenazi, M. (1999). Using Automatic Speech Processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning and Technology*, 2 (2): 62–76.

Flege, J. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues,* 233–277. Baltimore, MD: York Press.

Flege, J. (1999). The relation between L2 production and perception. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (eds), *Proceedings of the XIV International Congress of the Phonetic Sciences*, Vol. 2, 1273–1276. Berkeley, CA: University of California.

Flege, J., Takagi, N. and Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of Acoustical Society of America,* 99 (2): 1161–1173. http://dx.doi.org/10.1121/1.414884

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Godwin-Jones, R. (2009). Emerging technologies: personal learning environments. *Language Learning and Technology,* 13 (2): 3–9.

Gottfried, T. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, 12: 91–114.

Hale, M. and Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, 29: 656–683. http://dx.doi.org/10.1162/002438998553914

Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning?, *Speech Communication*, 51 (10): 906–919. http://dx.doi.org/10.1016/j.specom.2008.12.004

Hardison, D. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8 (1): 34–52.

Hardison, D. (2005). Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal* 22 (2): 175–190.

Hattori, K. (2009). *Perception and Production of English /r/-/l/ by Adult Japanese Speakers*. Unpublished doctoral dissertation. University College London, UK.

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL*, 15, 3–20. http://dx.doi.org/10.1017/S0958344003000211

Holec, H. (1981). *Autonomy and Foreign Language Learning*. Oxford: Pergamon.

Holland, M. (1999). Tutors that listen. *CALICO Journal*, 16 (3): 245–250.

Jenkins, J. (2000). *The Phonology of English as an International Language: New Models, New Norms, New Goals.* Oxford: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23 (1): 83–103. http://dx.doi.org/10.1093/applin/23.1.83

Jongman, A. and Wade, T. (2007). Acoustic variability and perceptual learning: The case of non-native accented speech. In O.-S. Bohn and M. J. Munro (eds), *Language Experience in Second Language Speech Learning*, 135–150, Amsterdam: John Benjamins.

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

Jurafsky, D. and Martin, A. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd Edition. Upper Saddle River, NJ: Prentice Hall.

Kawai, G. and Hirose, K. (2000). Teaching the pronunciation of Japanese double-mora phonemes using speech recognition technology. *Speech Communication*, 30 (2–3): 131–143. http://dx.doi.org/10.1016/S0167-6393(99)00041-2

Kennedy, C. and Levy, M. (2008). L'italiano al telefonino: Using SMS to support beginners' language learning. *ReCALL*, 20 (3): 315–330. http://dx.doi.org/10.1017/S0958344008000530

Kiernan, P. and Aizawa, K. (2004). Cell phones in task based learning. Are cell phones useful language learning tools? *ReCALL*, 16 (1): 71–84. http://dx.doi.org/10.1017/S0958344004000618

Kim, I. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology and Society*, 9 (1): 322–344.

King, R. (1967). Functional load and sound change. Language, 43 (4): 831–852. http://dx.doi.org/10.2307/411969

Koerich, R. (2006). Perception and Production of vowel paragorge by Brazilian EFL students. In B. Baptista and M. Watkins (eds), *English with a Latin Beat. Studies in Portuguese/Spanish – English Interphonology,* 91–104). Studies in Bilingualism 31. Amsterdam: John Benjamins.

Kukulska-Hulme, A. and Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, 20 (3): 271–289. http://dx.doi.org/10.1017/S0958344008000335

LaRocca, S., Morgan, J. and Bellinger, S. (1999). On the path to 2X learning: Exploring the possibilities of advanced speech recognition, *CALICO Journal* 16 (3): 295–310.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27: 1–19. http://dx.doi.org/10.1017/S0267190508070098

Levy, E. and Law II, F. (2010). Production of French vowels by American-English learn-

ers of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America*, 128 (3): 1290–1305. http://dx.doi.org/10.1121/1.3466879

Levy, E. and Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36 (1): 141–157. http://dx.doi.org/10.1016/j.wocn.2007.03.001

Littlewood, W. (2004). The task-based approach: Some questions and suggestions. *English Language Teaching Journal*, 58 (4): 319–326. http://dx.doi.org/10.1093/elt/58.4.319

Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning*, 24 (6): 515–525. http://dx.doi.org/10.1111/j.1365-2729.2008.00289.x

MacDonald, D., Yule, G. and Powers, M. (1994) Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning*, 44 (1): 75–100. http://dx.doi.org/10.1111/j.1467-1770.1994.tb01449.x

Mostow, J. and Aist, G. (1999). Giving help and praise in a reading tutor with imperfect listening because automated speech recognition means never being able to say you're certain. *CALICO Journal* 16 (3): 407–424.

Neri, A., Cucchiarini, C. and Strik, H. (2003). Automatic speech recognition for second language learning: how and why it actually works. *Proceedings of 15th International Congress of Phonetic Sciences*, 1157–1160, Barcelona, Spain.

Neri, A., Cucchiarini, C. and Strik, H. (2006). Selecting segmental errors in L2 Dutch for optimal pronunciation training. *International Review of Applied Linguistics*, 44 (4): 357–404. http://dx.doi.org/10.1515/IRAL.2006.016

Neri, A., Mich, O., Gerosa, M. and Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21 (5): 393–408. http://dx.doi.org/10.1080/09588220802447651

Nikolova, O. (2002). Effects of students' participation in authoring of multimedia materials on student acquisition of vocabulary. *Language Learning and Technology* 6 (1): 100–122.

Nunan, D. (2004). *Task-Based Language Teaching*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511667336

Rabiner, L. and Juang, B. (1993). *Fundamentals of Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.

Rochet, B. (1995). Perception and production of Second-Language speech sounds by adults. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, 379–410. Timonium, MD: York Press.

Rosa, E. and Leow, R. (2004). Computerized task-based exposure, explicitness, type of feedback, and Spanish L2 development. *Modern Language Journal*, 88 (2): 192–216. http://dx.doi.org/10.1111/j.0026-7902.2004.00225.x

Rosen, K. and Yampolsky, S. (2000). Automatic speech recognition and a review of its functioning with dysarthric speech. *Augmentative and Alternative Communication*, 16 (1): 48–60. http://dx.doi.org/10.1080/07434610012331278904

Schwienhorst, K. (2008). *Learner Autonomy and CALL Environments*. New York: Routledge.

Sheldon, A. (1985). The relationship between production and perception of the /r/–/l/ contrast in Korean adults learning English: A reply to Borden, Gerber, and Milsark. *Language Learning*, 35 (1): 107–13. http://dx.doi.org/10.1111/j.1467-1770.1985.tb01018.x

Sheldon, A. and Strange, W. (1982). The Acquisition of /r/-/l/ by Japanese Learners of English: Evidence that Speech Production Can Precede Speech Perception. *Applied Psycholinguistics*, 3 (3): 243–261. http://dx.doi.org/10.1017/S0142716400001417

Stampe, D. (1973). *A Dissertation in Natural Phonology*. New York: Garland.

Strambi, A. (2001). *The interaction of web-based interaction and collaboration on the language learner*. Unpublished doctoral thesis, University of Sydney.

Strik, H., Truong, K., Wet, F. and Cucchiarini, C. (2009). Comparing different approaches for automatic pronunciation error detection, *Speech Communication*, 51 (10): 845–852. http://dx.doi.org/10.1016/j.specom.2009.05.007

Warschauer, M. (1996). Comparing face-to-face and electronic communication in the second language classroom. *CALICO Journal* 13 (2): 7–26.

Young, V. and Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and the implications for speech-based applications used by the elderly: a literature review. *Assistive Technology Journal*, 22 (2): 99–112. http://dx.doi.org/10.1080/10400435.2010.483646

Zhang, H., Song, W. and Burston, J. (2011). Reexamining the effectiveness of vocabulary learning via mobile phones. *The Turkish Online Journal of Educational Technology,* 10 (3): 203–221.

Appendix. Sample of pronunciation form used in weekly activities by the ASR Group

Using Speech Recognition Week 1 Name_____

Pronunciation form

(Please, spend 1 minute per word/expression)

Word	# of attempts	Succeeded? (Yes/No)	if "No", what word(s) do you see on the screen?
radio			
pour			
tu es grand			
lire			
amour			
une vie agréable			
pur / pure			
cours			
maman			
tour			
partir			
lecture			
nous avons pu			
une grande table			
manger du poulet			
cours de français			
un peu			
il est sûr			
courir			
durée			
partout			