ReCALL **27**(3): 321–339. 2014 © European Association for Computer Assisted Language Learning
doi:10.1017/S0958344014000421
First published online 30 December 2014

Computer-mediated input, output and feedback in the development of L2 word recognition from speech

JOSHUA MATTHEWS

School of Education, University of Newcastle, Callaghan, Australia
(email: joshua.matthews@uon.edu.au)

JUNYU CHENG

School of Foreign Languages, Southeast University, 2 Si Pai Lou, Nanjing 210096, P. R. China;
(email: chjy@seu.edu.cn)

JOHN MITCHELL O'TOOLE

School of Education, University of Newcastle, Callaghan, Australia;
(email: mitch.otoole@newcastle.edu.au)

Abstract

This paper reports on the impact of computer-mediated input, output and feedback on the development of second language (L2) word recognition from speech (WRS). A quasi-experimental pre-test/treatment/post-test research design was used involving three intact tertiary level English as a Second Language (ESL) classes. Classes were either assigned to a control group ($n = 31$) or to one of two alternative treatment levels which used a web-based computer application enabling self-determined opportunities to repeatedly listen to and reconstruct spoken target text into its written form. Treatment group one ($n = 30$) received text feedback after each of their efforts at target text reconstruction, whereas treatment group two ($n = 35$) did not. Results indicated that word recognition gain scores of those who used the application, regardless of treatment level, were significantly higher than those of the control group. The relationship between the quantity of self-determined exposure to input and word recognition improvements was moderate but not linear, with those choosing moderate levels of speech input deriving the greatest measurable improvement. Neither increased levels of modified output nor the provision of text feedback were associated with significant improvements in word recognition gain scores. Implications for computer-mediated approaches for the development of L2 WRS are described and areas for future empirical research are suggested.

Keywords: L2 word recognition from speech, listening, CALL design, feedback, input, output

1 Introduction

Word recognition from speech (WRS) is the most important component of spoken language processing and the foundation of listening comprehension (Broersma & Cutler, 2008; Rost, 2002). This is the case as mapping elements of the speech signal onto specific lexical types held in the mental lexicon enables semantic word knowledge to be reliably accessed and put to use in the listening comprehension process (McQueen, 2007). The degree to which this process occurs with automaticity is particularly important for language learners: unlike written words on a page, spoken language is "distributed in time and fades quickly from the perceptual field" (Weber & Scharenborg, 2012: 387). This attribute makes word recognition from speech a highly time-constrained and implicit process which doesn't lend itself to the application of explicit language knowledge (Hulstijn, 2003). Consequentially, L2 WRS is an aspect of language learning which L2 learners find particularly difficult (Goh, 2000). A consideration of how to help learners deal with this difficulty is pertinent to language educators as lower level listening skills differentiate the more and less skilled L2 listeners (Tsui & Fullilove, 1989). Despite this importance, research focusing on L2 WRS is scarce (Broersma & Cutler, 2008). The research presented here seeks to tackle the research problem of how to build increased capacity of learners to recognise L2 words from speech in a real language learning context. In an effort to achieve this objective, a computer application specifically developed to improve L2 WRS was designed and evaluated using the tripartite framework of input, output and feedback.

1.1 The role of input, output and feedback in developing word recognition from speech

The constructs of input, output and feedback figure prominently in second language acquisition theory (Ellis, Loewen, Elder, Erlam, Philp & Reinders, 2009; Gass & Mackey, 2006; Krashen, 1982, 1985; McDonough, 2005; Swain, 1985), and also align with previously untested pedagogical recommendations on how to improve L2 WRS (Field, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). Although the themes of input, output and feedback have been applied to previous theoretical and practical research in the field of computer assisted language learning (CALL) (Chapelle, 2009; Heift, 2004; Li & Hegelheimer, 2013; Murphy, 2010), their role in the development of computer-mediated L2 WRS has not been rigorously explored in language learning contexts.

The role of frequency of input has a prominent position in theories which seek to explain the acquisition of fluent language skills. The power law of practice asserts that the speed and accuracy of skills such as word recognition from speech, increase in step with the frequency of skill-specific processing events (N. Ellis, 2002; Hulstijn, 2002). Providing opportunities for learners to repeatedly listen to spoken input is suggested to assist in developing the ability to recognise the location of boundaries between words and to improve automaticity of word recognition (Field, 2008a; Hulstijn, 2003). Previous studies show that the frequency of occurrence of words in aural input has a positive correlation with vocabulary learning, but that this variable is only one of several which are likely to contribute to the efficacy of L2 word learning (Nation, 2001; Vidal, 2003).

The production of output also plays an important role in language acquisition (Swain, 1985, 1995, 2000). According to Swain (1993: 159), "producing language forces learners to recognize what they do not know or partially know". In the case of L2 WRS, Field (2008a) recommends students repeatedly listen to fluent speech and attempt to accurately transcribe lexical content while critically reflecting on perceptual evidence and how it relates to tentative interpretations. Presumably this mode of output production pushes learners to reflect carefully on the exact lexical content of spoken language while also providing an explicit representation of the learner's perception of the spoken form.

Providing feedback which helps learners to notice deficits in their ability to recognise the lexical content of target language may also improve L2 WRS (Hulstijn, 2003; Wilson, 2003). Evidence suggests that "orthographic information can, under certain circumstances, have a significant impact on the long-term implicit, and explicit, learning of spoken word forms" (Bird & Williams, 2002: 529). Providing written feedback enables learners to engage explicit knowledge, and thus consciously attend to problematic elements of speech, and has practical benefits as the fixed form of written words is more readily accessible to learners in a conscious and explicit manner. Further, the role of explicit instruction has been comprehensively reviewed as an important element of effective L2 instruction (Norris & Ortega, 2000). However, the degree to which explicit knowledge, such as that which is engaged during the production of written target output or when attending to feedback, will facilitate improvements in language skills which are highly dependent on fluent performance, remains an issue of considerable debate (Ellis, 2005; Hulstijn, 2005).

Input, output and feedback have been suggested as a useful frame for the implementation of practical interventions aimed at the development of L2 WRS. This research will address the role of these constructs as design features of a web application specifically designed for this purpose.

1.2 CALL for the development of L2 word recognition from speech

Despite assertions that digital technology is well suited for the development of lower level listening skills (Vandergrift, 2007), research with the core objective of using CALL for the development of L2 WRS is scarce. The work of Hulstijn (2003), an influential exception, describes software specifically designed to improve L2 WRS. The software Hulstijn describes enables learners to listen repeatedly, attempt to reconstruct text and then verify predictions of lexical content using the written text transcript. Notwithstanding an absence of published empirical data verifying the effectiveness of these recommendations in real language learning contexts, Hulstijn's (2003) recommendations have been cited by leading authors in the fields of CALL (Chapelle & Jamieson, 2008; Levy, 2009) and L2 listening (Vandergrift, 2007) as a model of practice for the development of L2 WRS.

Matthews and O'Toole (2013) developed a web-based computer application designed to improve L2 WRS based on an elaboration of Hulstijn's recommendations, and investigated its use by a cohort of 33 ESL learners. The application enabled learners to listen multiple times, reconstruct target spoken language in writing in a number of phases and receive on-screen feedback highlighting discrepancies between the learners' output and the target language. The results were suggestive of the effectiveness of computer-mediated approaches which facilitate increased levels of exposure to input, increased production of output and the delivery of feedback. However, the study's one-group design did not allow the word recognition improvements observed among participants to be directly attributed to learning effects associated with their use of the application. Furthermore, the design of the study did not enable a refined investigation of the relative impact of input, output and feedback on L2 WRS development.

2 Research questions

The current study seeks to expand on the few previous theoretical and empirical investigations which have explored the role of CALL in the development of L2 WRS by addressing the following research questions:

1. What is the impact of the computer-mediated approach described here in the development of L2 word recognition from speech?
2. What is the impact of computer-mediated text feedback on L2 word recognition from speech?
3. What is the relationship between computer-mediated exposure to input and L2 word recognition from speech?
4. What is the relationship between computer-mediated production of modified output and L2 word recognition from speech?

3 Methodology

3.1 Participants

Three first-year undergraduate classes enrolled in the same general English language course, studying at a Chinese university, were involved in the study. Each class was kept intact, but randomly assigned to one of three levels: a control group ($n = 31$), treatment level one ($n = 30$) or treatment level two ($n = 35$). All students spoke Mandarin Chinese as a first language and were aged between 17 and 20. Objectives of the course prescribed by the cooperating institution included the development of listening ability and the use of computers to improve language learning outcomes. All participants were accustomed to regular computer and internet usage.

3.2 Materials for the computer application

The listening materials used with the application contained factual information about countries in the Asian region based on piloted material. Eight short monologues, with an average of 67 words were recorded in MP3 format. The lexical content of the monologues was analysed using Vocabprofile (Cobb, 2014; Heatley & Nation, 1994) in order to determine the proportions of high frequency, academic, and low frequency words (Coxhead, 2000; Nation, 2001; West, 1953). The relative proportions of these word categories in the monologues were comparable to those found in fiction, newspapers, and academic texts (Nation, 2001). The average speech rate of the monologues was 2.1 words per second, which approximates the speech rate of an academic lecture presented at average speed (Tauroza & Allison, 1990). Monologues were edited into shorter monologue sections, with an average length of 7.3 words, such that their average duration approximated that of the intonation units typical of fluent speech (Rost, 2002). For each monologue, a static contextual image was sourced in order to be used as a pre-task advanced organiser. The speech input from each monologue section was also transcribed into a text file for use with the application. The combined use of monologues, monologue sections, static images and transcript texts will be discussed later in this paper.

3.3 Overview of the application

The desired design elements for the web application were specified and collaboratively developed by the primary author and a small team of professional web-application

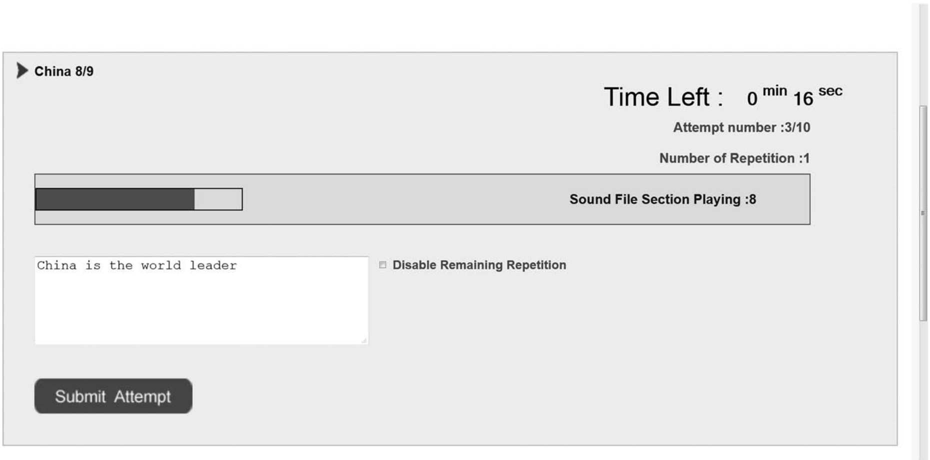

Fig. 1. Opportunities to increase exposure to input and to produce modified output

developers. Open source web server software and database management systems were used. The application was deployed on a remote web server.

Access via the administrator interface allowed selection of the sequence and timing of any number of sound files, text files and image files to be established prior to user access. Key variables which were set in relation to the current study were the number of maximum opportunities learners had to listen to target language, the opportunity to enter text in response to hearing sound files, the presence or absence of text feedback, and the option to allow learners freedom to proceed to the next sound file at their own pace. Key data captured by the database were the number of times users chose to listen to each sound file, the text output produced by users in response to each time a monologue section was heard, and the feedback mode of each participant.

The web application first played the complete monologue while the contextual static image remained on-screen. After listening to the full monologue, participants listened to the nine monologue sections one after the other. Clicking an on-screen panel enabled the participants to control when the monologue section began playing, after which the monologue section played all the way through without pauses. Participants were able to listen to each monologue section between one and ten times and were instructed to choose to listen as many times as they felt was appropriate for their learning needs. This design feature enabled participants to increase their exposure to target speech input in a manner which students themselves deemed to be most appropriate (Figure 1).

Each time the monologue section was played, an opportunity to transcribe the aural text by typing reconstructions on-screen was provided. Each time participants chose to listen to the monologue section again, their previous effort at transcription was maintained on-screen such that it could be edited. This design feature enabled the production of output in written form. It also provided an opportunity to recast efforts at transcription in response to the perceptual evidence available from the speech input (Figure 2). As mentioned previously, each effort at transcription was recorded in the data base, enabling quantification of the number of times written output was modified by each participant for each monologue section. The construct of modified output was determined as the number of times

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

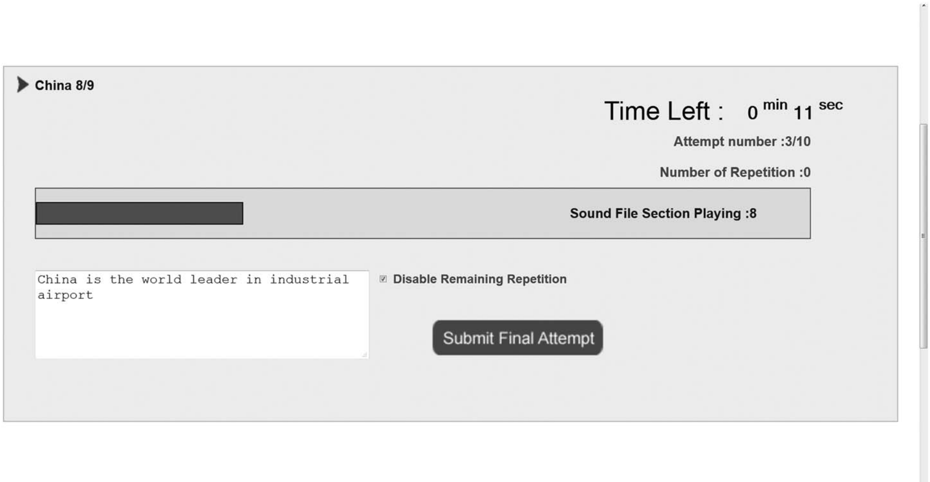

Fig. 2. Opportunities to increase exposure to input and to produce modified output

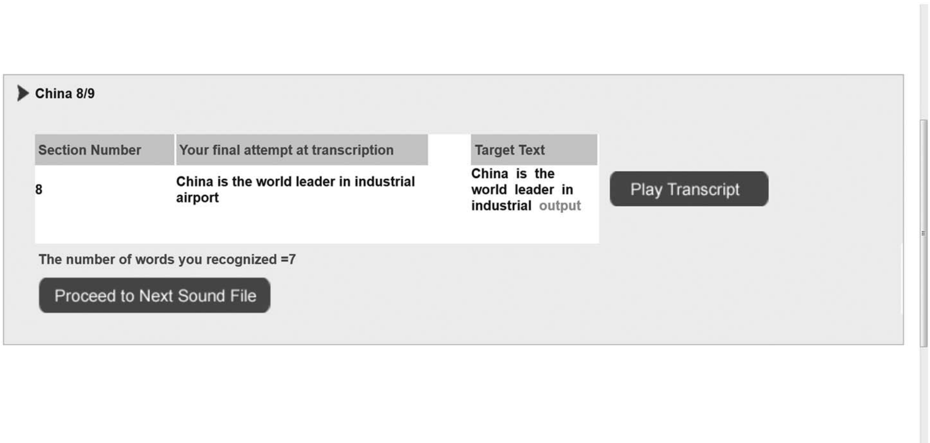

Fig. 3. Example of feedback mode of treatment level 1 participants (text feedback provided)

participants altered their written attempts at text reconstruction for each monologue section. These data were extracted from participants' text output and coded manually, then checked digitally with a coded algorithm.

After listening to the monologue section the chosen number of times, one of the two feedback formats was made available. Treatment level one participants received a numerical indication of the number of words they correctly transcribed as well as the target text with errors or omissions highlighted in red (Figure 3). Treatment level two participants received feedback relating to the number of words correctly transcribed but did not receive target text feedback (Figure 4). Participants at both treatment levels were provided the opportunity to listen to the monologue section once again by clicking the "Play transcript" button.

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

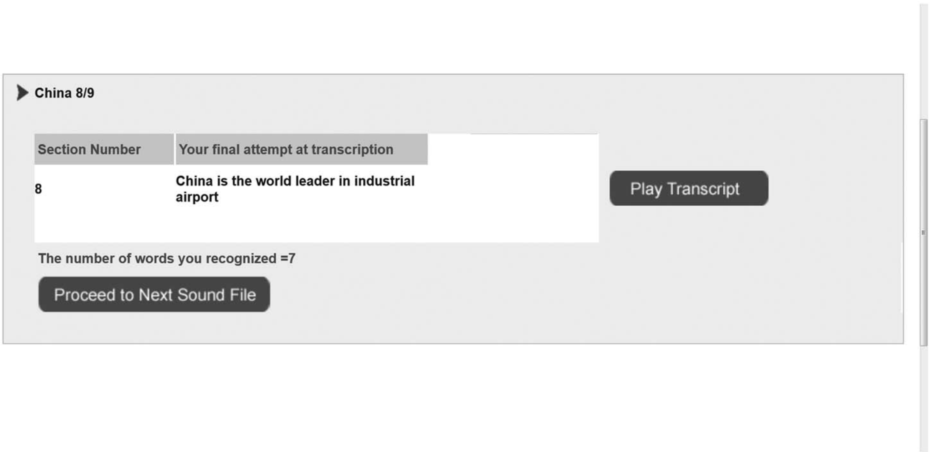

Fig. 4. Example of feedback mode of treatment level 2 participants (no text feedback provided)

4 Test instruments

4.1 Receptive vocabulary knowledge test

A measure of the receptive vocabulary knowledge of the participants was determined using the second thousand-word level and the academic word level components of a version of the Vocabulary Levels Test (Nation, 1983, 1990; Schmitt, Schmitt & Clapham, 2001). Words from the second thousand-word level are those which are traditionally considered to be high frequency vocabulary items. Academic words are those which occur beyond the second thousand-word level and which occur relatively frequently in a variety of academic texts (Nation, 2001). The portion of the test used consisted of 66 target words with 30 high-frequency and 36 academic target words. As this test format has been validated in a number of contexts (Beglar & Hunt, 1999; Schmitt *et al*., 2001) it was used as the criterion for concurrent validation of the word recognition pre- and post-tests described below.

4.2 Word recognition tests

4.2.1 Test structure and specifications. Word recognition was measured with two partial dictation tests specifically designed for the purposes of this study (pre-test, see Appendix 1; post-test, see Appendix 2). These tests involve test-takers completing a contextual sentence by listening to a spoken stimulus and filling in the blank. Each partial dictation test (hereafter *word recognition test*) consisted of 60 items: the first 32 target words were high frequency words, while the last 28 were from the Academic Word List (Coxhead, 2000; Nation, 2001). The target words for odd-numbered items for both the pre-test and the post-test were words present within the content of the monologues used in the learning intervention (hereafter *intervention target words*). Target words for even-numbered items for both the pre-test and the post-test were words which did not appear in the learning intervention monologues (hereafter *general target words*). Therefore the word recognition pre-test and post-test consisted of sixteen high frequency intervention words, sixteen high frequency general words, fourteen academic intervention words and fourteen academic

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

general words. The intervention target words of the pre-test and post-test were the same, but to minimise potential learning effects from one test administration to the next, each word was embedded in a different contextual sentence. The general target words of the pre-test and the post-test were different.

4.2.2 Construct and concurrent test validation. A two-part process was used to ensure the primary construct being measured by the word recognition tests was the ability to recognise the phonological form of the target words. First, a panel of adult native English language speakers completed the word recognition test items without listening to the spoken stimulus. This was to ensure that the target words could not be systematically recognised by using contextual cues present within the written sentences of the items. Second, a new panel of native speakers undertook the tests, but this time with the accompanying spoken language stimulus. All panel members recognised all target words correctly. This process ensured that the contextual information in the item stem was not the primary source of information used to recognise the target word and that the WRS capacity required for high levels of test achievement did not exceed that of expert target language users.

The participants' receptive vocabulary knowledge scores (criterion) were strongly correlated with both word recognition pre-test scores, $r = 0.73$, $N = 96$, $p < 0.01$ and post-test scores, $r = 0.74$, $N = 96$, $p < 0.01$.

4.2.3 Word recognition test scoring procedures. Word recognition scores were assigned by using a structured scoring rubric which limited the influence of spelling errors on scoring (Buck, 2001). This was achieved by assigning a mark to target words written in the correct orthographic form or in a form that was clearly recognisable despite minor spelling errors. Words which were written with a degree of ambiguity, but which were still recognisable as the target word, were ascribed a half mark. In order to standardise this procedure, the scoring rubric was used by two trained, independent native speaking test scorers. Cohen's kappa analysis indicated a strong level of inter-rater agreement ($\kappa = 0.91$, $p < 0.01$). Any discrepancies between the scorers were resolved by consultation between them with reference to the scoring rubric after initial scoring had been completed.

4.2.4 Test reliability. A split halves reliability analysis was used to quantify internal consistency of the word recognition pre-test and the post-test. Scores achieved for odd and even test items were first correlated and then corrected with the Spearman-Brown formula. Both tests were found to have a high level of internal consistency (Table 1).

Reliability coefficients between pre- and post-tests were determined through bivariate correlational analysis. Word recognition pre- and post-test scores were strongly and positively correlated; $r = 0.89$, $N = 96$, $p < 0.01$.

Table 1 *Summary of internal consistency measures for pre-test and post-tests*

Test and scoring protocol	Correlation between split halves forms	Spearman-Brown coefficient
Pre-test: partial word recognition	0.88	0.94
Post-test: partial word recognition	0.89	0.94

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

5 Procedure

The study was undertaken within a seven-week period, with the first and last weeks allocated to testing. All students ($N = 96$) undertook the receptive vocabulary knowledge test and the word recognition pre-test in the first week of the study. Participants in the treatment groups were provided with log-on credentials and a short in-class demonstration of the use of the computer application. Each of these students accessed the computer application via the internet with an individual computer and listened through headphones. Five 60-minute sessions took place in a computer laboratory once a week for five weeks. All 65 treatment group participants completed the eight monologues within the allocated 300-minute time frame. Word recognition post-tests were administered to all 96 participants in the seventh week of the research period. All three groups were taught by the same language instructor and received similar learning activities other than those undertaken by the two treatment groups once a week in the computer laboratories.

6 Analysis

In order to gauge improvements in L2 WRS, two gain scores were calculated for each participant:

1. Word recognition gain score for intervention target words.
2. Word recognition gain score for general target words.

Gain scores were calculated by subtracting the appropriate post-test score from the corresponding pre-test score. Research questions were primarily investigated by comparing mean values of these two gain scores grouped in the following ways:

1. Participants grouped based on whether they used the application or not.
2. Participants grouped based on whether they belonged to the control group, treatment group one (text feedback) or treatment group two (no text feedback).
3. Participants grouped on relative exposure to input (treatment group only).
4. Participants grouped on relative production of modified output (treatment group only).

The statistical analyses used to compare the mean gain scores for these groupings were either independent samples *t*-test or one-way analysis of variance (ANOVA). Shapiro-Wilk tests, coupled with visual appraisal of gain score histograms and normal probability plots, indicated that these data were normally distributed. Levene's tests confirmed homogeneity of variance among all sets of group means compared in the analyses.

7 Results

7.1 Research question one

What is the impact of the computer-mediated approach described here in the development of L2 word recognition from speech?

In order to investigate the impact of using the web application on L2 WRS, participants' gain scores from both treatment groups were combined to form a single group and mean gain scores were calculated. Independent samples t-tests were conducted to compare the mean WRS gain scores attained by those participants who used the application (treatment

Table 2 *Mean WRS gain scores for participants who used the application and those who did not*

Category of target word	Group	Mean word gain score (words)	*SD*	Sig. (2-tailed)	Effect size (Cohen's *d*)
Intervention target words	Treatment	6.44	3.26	0.04*	0.47
	Control	5.03	2.71		
General target words	Treatment	4.93	3.26	0.45	0.16
	Control	4.39	3.37		

*$p < 0.05$

Table 3 *Mean WRS gain scores for treatment one, treatment two and control group*

Category of target word	Group level	Mean word gain score (words)	*SD*	Sig. (2-tailed)	Effect size (η^2)
Intervention target words	Treatment 1	6.77	3.89	0.09	0.05
	Treatment 2	6.16	2.64		
	Control	5.03	2.71		
General target words	Treatment 1	4.35	2.50	0.32	0.02
	Treatment 2	5.42	3.76		
	Control	4.39	3.37		

groups one and two) and those participants in the control group who did not use the application (Table 2).

Intervention word recognition gain scores were significantly greater for those participants who used the application ($M = 6.44$, $SD = 3.26$) than for participants in the control group ($M = 5.03$, $SD = 2.71$); $t\ (94) = 2.08$, $p < 0.05$. No significant differences were observed between the two groups in relation to general target word gain scores. Although these results suggest that the use of the application, either with or without text feedback, was associated with significant improvements in L2 WRS, these positive effects were confined to intervention target words.

7.2 Research question two

What is the impact of computer-mediated text feedback on L2 word recognition from speech?

In order to quantify the impact of text feedback on L2 WRS, mean word recognition scores for each of the three groups (treatment group one, treatment group two and control) were compared using a one-way ANOVA. The main focus of this analysis was the comparison of mean word recognition gain scores for treatment group one and treatment group two. However, control group mean gain scores were also of interest as they indicate the levels of word recognition development likely to have occurred for various reasons not directly linked to the use of the application. No significant differences in mean gain scores in word recognition for either intervention or general target words were observed (Table 3).

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

Table 4 *Mean WRS gain scores for low, moderate and high exposure to input categories*

Category of target word	Relative exposure to input category	Mean word gain score (words)	*SD*	Sig. (2-tailed)	Effect size (η^2)
Intervention target words	Low	5.11	2.36	0.02*	0.12
	Moderate	7.86	3.81		
	High	6.33	2.97		
General target words	Low	5.50	2.94	0.15	0.06
	Moderate	5.45	2.80		
	High	3.79	3.84		

*$p < 0.05$

7.3 Research question three

What is the relationship between computer-mediated exposure to input and L2 word recognition from speech?

To determine if differences in the amount of exposure to speech input had a measurable impact on WRS gain scores, treatment group one and treatment group two were pooled into a single group ($n = 65$). Participants from this combined group were then allocated to either low, moderate or high exposure categories based on 33 percentile cut-off points for the mean number of times each participant chose to listen to the 72 monologue sections. The mean number of times each monologue section was listened to by the entire group was 4.44 times ($SD = 1.22$, min. = 2.46, max. = 8.61). The low exposure category had fewer than 3.82 repetitions per monologue section; moderate exposure participants had on average between 3.82 and 4.72 repetitions per section; high exposure category participants had over 4.72 repetitions per monologue section. These categories were used as the grouping variables for analysis of differences between mean WRS gain scores (Table 4).

A one-way ANOVA showed a significant impact of relative exposure to input on WRS gain scores for intervention target words, $F(2,62) = 4.33$, $p < 0.05$, $\eta^2 = 0.12$. Tukey's Honest Significant Difference (HSD) tests indicated that only differences between gain scores for low ($M = 5.11$, $SD = 2.36$) and moderate ($M = 7.86$, $SD = 3.81$) exposure categories were significant ($p < 0.05$). A one-way ANOVA showed no significant impact of relative exposure to input on recognition of general target words.

In summary, the amount of self-determined exposure to input had a statistically significant association with the differences in mean WRS gain scores observed between groups. However, this relationship was confined to improvements on intervention words and to differences observed between participants from low and moderate categories of input exposure. A Pearson product-moment correlation coefficient indicated a moderate positive relationship between mean exposure to input and word recognition gain scores among low and moderate exposure category participants only ($r = 0.35$, $n = 44$, $p < 0.05$). Correlational analysis of a combined group consisting of participants from all three categories, including the high exposure group, yielded an explanatory model which was clearly non-linear. These results suggest that increased levels of exposure to input were associated with WRS improvements up to a point, beyond which the positive relationship between exposure to input and word recognition improvements for intervention target words ceased.

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

Table 5 *Mean WRS gain scores for low, moderate and high modified output categories*

Category of target word	Relative production of modified output	Mean word gain score (words)	*SD*	Sig. (2-tailed)	Effect size (η^2)
Intervention target words	low	6.13	3.01	0.66	0.01
	moderate	6.22	3.58		
	high	6.98	3.21		
General target words	low	5.58	3.19	0.57	0.02
	moderate	4.69	2.13		
	high	4.60	4.12		

7.4 Research question four

What is the relationship between computer-mediated production of modified output and L2 word recognition from speech?

Again, treatment group one and treatment group two were pooled into a single group ($n = 65$) and categorised according to relative levels of modified output productivity. The mean number of times participants modified their efforts at text reconstruction for the 72 monologue sections was 2.23 times per section ($SD = 0.62$, min. = 1.00, max. = 4.04). The low output category participants had mean modified output values of less than 1.93 modifications per monologue section, the moderate output category between 1.93 and 2.39 modifications, and the high category more than 2.39 modifications. These categories were used as the grouping variables for the ANOVA in order to analyse differences in mean WRS gain score between the groups (Table 5).

A one-way ANOVA showed no significant relationship between mean modified output categories on either intervention or target WRS gain scores. The small effect sizes shown in Table 5 confirm the limited practical impact of the relative output modification category on differences in mean WRS gain scores.

8 Discussion and conclusions

The finding that those participants who used the application experienced significantly greater improvements in L2 WRS than those in the control group provides empirical support for the value of CALL in the development of L2 WRS. These results suggest that similar interventions, applied among those with similar linguistic attributes to those involved in this study, are likely to yield positive L2 WRS improvements within relatively short periods of time. However, if the results of this research do indeed provide a useful predictive model for other learning contexts, then any such word recognition improvements are likely to be confined to those words which are present in the speech input used as part of the learning intervention. These findings provide insight into the specificity of such learning approaches and encourage strategic pedagogical decisions about the listening material used for the purposes of developing L2 WRS. Specifically, it is important that the lexical content of listening materials used clearly matches the specific word learning objectives of the language courses or groups of students in question. Ensuring learners have the ability to recognise high frequency vocabulary and the specialised vocabulary particular to their field of anticipated expertise would be a logical and practical starting point.

Another significant finding was that a positive relationship between the quantity of exposure to input and WRS gain scores was only evident among participants in the low and moderate input level groups. Participants choosing to listen to high category levels of input did not significantly outperform those that listened to low and moderate category levels. There are several alternative explanations for the non-linear relationship observed between exposure to input and WRS gain scores. First, as suggested by previous research, increasing exposure to input is likely to only be of benefit to those learners who have the underlying linguistic proficiency necessary to derive its potential benefit (VanPatten, Williams & Rott, 2004; Vidal, 2011). In support of this assertion, a correlational analysis indicated that mean self-determined exposure to input was negatively correlated with word recognition pre-test scores ($r = -0.55$, $n = 65$, $p < 0.01$). This suggests that those participants who chose to listen to greater levels of input had relatively low baseline L2 WRS proficiency. The implication is that those participants who did derive word recognition improvement in association with increasing levels of input did so with relatively few repetitions (less than 4.72) and had relatively high baseline levels of word recognition capability. In short, the relative benefit of repeated listening for improving WRS is likely to be influenced by the existing word recognition proficiency status of the individual language learner.

Overall these findings suggest that a careful consideration of input, specifically the learner's current knowledge status in relation to target language material, is likely to be of great importance in effective and targeted interventions aimed at the improvement of L2 WRS. However, caution must be exercised in drawing conclusions regarding the potentially causative effect of increased levels of input on improvements in WRS among high proficiency level students. The amount of input received by participants was a function of participants' free choice and uncontrolled by the researchers. It is therefore possible that other factors, such as learner motivation, contributed to the relationship between exposure to input and WRS improvement and thus must also be amply considered in the delivery of future L2 WRS interventions and research.

In contrast to the findings in relation to the role of input, the provision of text feedback was not associated with statistically significant improvements in L2 WRS. Similarly, increased levels of output through transcription of target speech into the written form were not associated with greater levels of WRS gains. The impact of differing levels of these constructs, analysed as isolated independent variables, was not sufficient to elicit statistically significant improvements in WRS over the time frame of the intervention. Contextualised within the framework of this study, these findings cast some doubt over the pedagogical value of explicit feedback in the development of L2 WRS. Although speculative, it may be the case that word knowledge developed through relatively explicit modes of learning is not readily transferrable to highly time constrained language processes such as those which underpin the ability to recognise words from speech. These findings are in line with previous research which shows word knowledge developed through explicit instruction is relatively unstable when applied to spoken word forms (Takimoto, 2009). Additionally, it may the case that measurable improvements in word recognition linked to the provision of text feedback or the opportunity to reconstruct written output require longer durations of time than those implemented as part of the current research. However, based solely on the results of this study, there is a lack of compelling quantitative evidence to support the role of text feedback and the production of output in the development of L2 WRS.

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

This study has endeavoured to provide additional empirical research relating to computer-assisted development of L2 WRS. However, there are several limitations of this study, the identification of which may provide a useful framework for future improvement and elaboration on the current work. First, although it was not feasible here, future research of a similar nature would be well served by randomly assigning individual participants to different grouping levels. Secondly, as previously mentioned, the timeframe of this study was relatively brief, involving just five weeks of intensive practice. Increasing the duration of similar future studies may yield learning trajectories discrepant from those summarised here. For example, it is reasonable to postulate that greater amounts of input over longer periods of time may been necessary for lower proficiency participants to derive measurable benefit from similar interventions (Vidal, 2011).

Additionally, for logistical reasons it was only possible to measure word recognition shortly after the completion of the last word recognition learning session. Delayed post-tests would provide additional information on the durability of the WRS improvements associated with use of this or similar applications. Further, this paper only reported on the quantitative results of the investigation; supplementing these with the qualitative information also gathered as part of this study in future reports will add additional insight to the findings presented here. Finally, investigating the comparative success of similar pedagogical approaches with different target languages and through use of mobile technology platforms would also be of interest.

References

Beglar, D. and Hunt, A. (1999) Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests. *Language Testing*, **16**(2): 131–162.

Bird, S. and Williams, J. (2002) The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, **23**(4): 509–533.

Broersma, M. and Cutler, A. (2008) Phantom word activation in L2. *System*, **36**(1): 22–34.

Buck, G. (2001) *Assessing listening*. Cambridge: Cambridge University Press.

Chapelle, C. (2009) The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal*, **93**: 741–753.

Chapelle, C. and Jamieson, J. (2008) *Tips for teaching with CALL*. White Plains, NY: Pearson Education.

Cobb, T. (2014) *Web Vocabprofile* [an adaptation of Heatley and Nation's (1994) *Range*]. http://www.lextutor.ca/vp

Coxhead, A. (2000) A new academic word list. *TESOL Quarterly*, **34**(2): 213–238.

Ellis, N. (2002) Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, **24**: 143–188.

Ellis, N. (2005) At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, **27**: 305–352.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J. and Reinders, H. (2009) *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.

Field, J. (2008a) Revising segmentation hypotheses in first and second language listening. *System*, **36**(1): 35–51.

Field, J. (2008b) *Listening in the language classroom*. Cambridge: Cambridge University Press.

Gass, S. and Mackey, A. (2006) Input, interaction and output: An overview. *AILA Review*, **19**: 3–17.

Goh, C. (2000) A cognitive perspective on language learners' listening comprehension problems. *System*, **28**(1): 55–75.

Heatley, A. and Nation, P. (1994) *Range* [Computer program]. http://www.vuw.ac.nz/lals

Heift, T. (2004) Corrective feedback and learner uptake in CALL. *ReCALL*, **16**(2): 416–431.

Hulstijn, J. (2002) Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, **18**(3): 193–223.

Hulstijn, J. (2003) Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning* **16**(5): 413–425.

Hulstijn, J. (2005) Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition*, **27**: 129–140.

Krashen, S. (1982) *Principles and practice in second language acquisition*. New York: Pergamon Press.

Krashen, S. (1985) *The input hypothesis: Issues and implications*. London: Longman.

Levy, M. (2009) Technology in use for second language learning. *The Modern Language Journal*, **93**: 769–782.

Li, Z. and Hegelheimer, V. (2013) Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language Learning & Technology*, **17**(3): 135–156.

Matthews, J. and O'Toole, J. M. (2013) Investigating an innovative computer application to improve L2 word recognition from speech. *Computer Assisted Language Learning*, 1–19. DOI: 10.1080/09588221.2013.864315.

McDonough, K. (2005) Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, **17**(1): 79–103.

McQueen, J. M. (2007) Eight questions about spoken word recognition. In: Gaskell, M. G. (ed.), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, 37–53.

Murphy, P. (2010) Web-based collaborative reading exercises for learners in remote locations: The effects of computer-mediated feedback and interaction via computer-mediated communication. *ReCALL*, **22**(2): 112–134.

Nation, I. S. P. (1983) Testing and teaching vocabulary. *Guidelines*, **5**: 12–25.

Nation, I. S. P. (1990) *Teaching and learning vocabulary*. Rowley, MA: Newbury House.

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

Norris, J. and Ortega, L. (2000) The effectiveness of L2 instruction: A research synthesis and qualitative meta-analysis. *Language Learning*, **50**(3): 417–528.

Rost, M. (2002) *Teaching and researching listening*. Harlow: Longman.

Schmitt, N., Schmitt, D. and Clapham, C. (2001) Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, **18**(1): 55–88.

Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In: Gass, S. and Madden, C. (eds.), *Input in second language acquisition*. Rowley, MA: Newbury House, 235–253.

Swain, M. (1993) The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, **50**: 158–164.

Swain, M. (1995) Three functions of output in second language learning. In: Cook, G. and Seidhofer, B. (eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson*. Oxford: Oxford University Press, 125–144.

Swain, M. (2000) The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In: Lantolf, J. (ed.), *Sociocultural theory and second language learning*. Oxford: Oxford University Press, 97–114.

Takimoto, M. (2007) The effects of input-based tasks on the development of learners' pragmatic proficiency. *Applied Linguistics*, **30**(1): 1–25.

Tauroza, S. and Allison, D. (1990) Speech rates in British English. *Applied Linguistics*, **11**(1): 90–105.

Tsui, A. and Fullilove, J. (1998) Bottom up or top down processing as a discriminator of L2 listening performance. *Applied Linguistics*, **19**(4): 432–451.

Vandergrift, L. (2007) Recent developments in second and foreign language listening comprehension research. *Language Teaching*, **40**(3): 191–210.

VanPatten, B., Williams, J. and Rott, S. (2004) Form-meaning connections in second language acquisition. In: VanPattern, B., Williams, J., Rott, S. and Overstreet, M. (eds.), *Form-meaning connections in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum, 1–26.

Vidal, K. (2003) Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, **24**(1): 56–89.

Vidal, K. (2011) A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, **61**(1): 219–258.

Weber, A. and Scharenborg, O. (2012) Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, **3**(3): 387–401.

West, M. (1953) *A General Service List of English Words.* London: Longman, Green and Co.

Wilson, M. (2003) Discovery listening: Improving perceptual processing. *ELT Journal*, **57**(4): 335–343.

APPENDICES

Appendix 1 – Word recognition pre-test

1. The most.. language is South Korean.
2. This country depends on..
3. .... 50 million people live in this country.
4. Parts of this country are covered in..
5. The city of Busan is in the..
6. Great.. has been put into improving this country.
7. This country.. on making cars.
8. There are many.. in this country.
9. The.. of this country is quite large.
10. Many.. Korea as a very beautiful country.
11. South Korea is a.. small country.
12. Korea doesn't have many..
13. Only one other country shares a.. with South Korea.
14. .... is not a large industry in this country.
15. The.. language of this country is Korean.
16. Daegu is the.. biggest city in South Korea.
17. ....is important to this country.
18. A product of this country is..
19. Korea.. business people from around the world.
20. Many people are.. by Korea's natural beauty.
21. The.. port in this country is in Busan.
22. Many people.. Korea every year.
23. .... about Korea can be found on the internet.
24. Korea has a very good.. system.
25. There are several.. schools in Korea.
26. Korean food is.. popular.
27. Korea has a large number of..
28. Most of this country is surrounded by..

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press

29. Korea produces..
30. The.. in this country changes very quickly.
31. This country doesn't produce natural..
32. The number of people in this country is.. growing.
33. Korea is a country with a strong..
34. Many.. work in South Korea.
35. It.. cars and computers.
36. Busan and Seoul are very.. cities.
37. The.. language is standard Korean.
38. This country has a good..system.
39. Recently this country has experienced.. growth.
40. Korea has many.. to improve production.
41. Korea has received.. from many different countries.
42. Korea needs to.. coal from other countries.
43. Milk is an important.. product of this country.
44. Korea has the.. to continue to modernise.
45. This country has many.. varieties of food.
46. Korea is a.. country.
47. China and Korea have strong..
48. Korea is a.. country in East Asia.
49. Korea's second largest city is.. in the south.
50. There are few.. in South Korea.
51. .... languages are uncommon in this country.
52. Korea has..with many countries.
53. Korea produces many types of..
54. Korea has developed a new trading..
55. In the past Korea has been..
56. .... sports are popular in Korea.
57. Recently the industrial.. of this country has increased.
58. Recently Korea has.. more trade with other countries.
59. The.. of Korean food makes it popular around the world.
60. Korea will.. several new laws next year.

Target words (in sequence)

common, shipping, currently, mountains, east, effort, depends, towns, population, regard, relatively, lakes, border, mining, standard, fourth, agriculture, fruit, attracts, charmed, busiest, tour, information, health, international, especially, islands, ocean, rice, weather, rubber, steadily, economy, contractors, exports, similar, major, financial, significant, projects, investment, source, primary, potential, regional, secure, links, dominant, located, immigrants, minority, partnerships, technology, scheme, occupied, professional, output, predicted, diversity, initiate

Appendix 2 – Word recognition post-test

1. Vietnamese is the most.. language.
2. The cost of living in this country is..
3. .... 90 million people live in Vietnam.
4. This country has many..

5. The city of Binh Dinh is in the..
6. Great ……………………………………. has been made to develop this country.
7. This country ………………………………….. on manufacturing.
8. There are many.. in this country.
9. The.. of this country is large.
10. Many.. Vietnam as a very beautiful country.
11. Vietnam is a.. long country.
12. An important food in Vietnam is..
13. Vietnam.. borders with three other countries.
14. .... is an important industry in this country.
15. Vietnamese is the.. language of this country.
16. Ba Ria is the.. largest city in Vietnam.
17. .... is very important to Vietnam.
18. One product of this country is..
19. Vietnam.. visitors from around the world.
20. Various types of.. are popular in Vietnam.
21. One of the.. cities in Vietnam is Hanoi.
22. .... are very important to this country.
23. .... about this country can be found on the internet.
24. This country has a good.. system.
25. This country has several.. schools.
26. Vietnamese food is.. popular.
27. This country has several famous..
28. On the coast of this country there is a large..
29. This country produces..
30. The.. in this country changes very quickly.
31. Vietnam produces a large amount of natural..
32. The number of people in this country is.. growing.
33. The.. of Vietnam is growing.
34. Vietnam produces a.. of things.
35. .... are very important to Vietnam.
36. Industry is the largest.. in this country.
37. Vietnamese is the.. language.
38. Within Vietnam there are many different..
39. This country has a.. number of different languages.
40. The.. in this country is very important.
41. Many foreign countries have made.. in this country.
42. Vietnam uses many.. to improve production.
43. Fish is an important.. product of this country.
44. This country has experienced a.. of market reforms.
45. Different parts of Vietnam have different.. languages.
46. Vietnam has many important.. sites.
47. Vietnam has.. with Cambodia.
48. Vietnam has experienced.. development.
49. Vietnam's second largest city is.. in the north.
50. This country is making a.. to Southeast Asian growth.

51. Many... languages exist in the country.
52. Vietnam has several oil...
53. .... is important to this country's development.
54. The protection of Vietnam's forests is an important...
55. This country has been... in the past.
56. .... travel is cheap in this country.
57. In recent times Vietnam's coal... has reduced.
58. Recently Vietnam has... trade with other countries.
59. There is a large... of languages and cultures.
60. There are various forms of... within this country.

Target words (in sequence)

common, rising, currently, forests, east progress, depends, villages, population, recognise, relatively, fish, shares, farming, standard, third, agriculture, coffee, attracts, dance, busiest, tourists, information, education, international, extremely, islands, bay, rice, temperature, rubber, gradually, economy, range, exports, sector, major, cultures, significant, environment, investments, processes, primary, series, regional, cultural, links, constant, located, contribution, minority, corporations, technology, task, occupied, domestic, output, promoted, diversity, transport

https://doi.org/10.1017/S0958344014000421 Published online by Cambridge University Press