L2 English Learners' Recognition of Words Spoken in Familiar versus Unfamiliar English Accents

Jia Ying[1, 2], *Jason A. Shaw*[1, 2], *Catherine T. Best*[1]

[1]The MARCS Institute, University of Western Sydney, Australia
[2]School of Humanities and Communication Arts, University of Western Sydney, Australia
16619366@student.uws.edu.au, j.shaw@uws.edu.au, c.best@uws.edu.au

Abstract

How do L2 learners cope with L2 accent variation? We developed predictions based upon the Perceptual Assimilation Model-L2 (PAM-L2) and tested them in an eye-tracking experiment using the visual world paradigm. L2-English learners in Australia with Chinese L1 were presented with words spoken in familiar Australian-accented English (AusE), and two unfamiliar accents: Jamaican Mesolect English (JaME) and Cockney-accented English (CknE). AusE and JaME differ primarily in vowel pronunciations, while CknE differs primarily in consonant pronunciations. Words were selected to elicit two types of perceptual assimilations of JaME and CknE phonemes to AusE: Category Goodness (CG) and Category Shifting (CS) assimilations. The Perceptual Assimilation Model (PAM) predicts that, if the L2 learners have developed AusE categories, then CS differences should hinder spoken word recognition more than CG differences. Our results supported this prediction. For both unfamiliar accents, CS target words attracted more fixations to printed competitor words than did CG distracters.

Index: cross-language speech perception, spoken word recognition, regional accent

1. Introduction

Language-specific experience determines how listeners handle natural variability in speech. Some recent studies show that even native speakers' reaction is slowed down and their accuracy rate is reduced when they hear a non-native regional accent [1]. For L2 learners, the challenge of regional accent variability is amplified, as many studies indicate, for example [2], [3], [4], and [5]. In this study, we attempt to pinpoint sources of difficulty for L2 learners in word recognition across accents. Using a visual world paradigm (see [6] and [7]), we investigated how Chinese learners of English familiar with Australian English process words produced in two English accents unfamiliar to them: Jamaican- and Cockney-accented English. Our predictions about which accent differences slow word recognition are derived from the Perceptual Assimilation Model-L2 (PAM-L2). PAM-L2 predicts that the phonological and phonetic relationship between L1 and L2 and language experience with L2 affect second language learners' perception of L2 phonemes [8]. According to Best [9], non-native speech sounds will be assimilated in one of three ways, and the type of assimilation will predict discrimination performance. The assimilation types most relevant to recognition of words spoken in an unfamiliar regional accent are Two Category assimilation (TC type) and Category Goodness assimilation (CG type). The TC case applies when a phoneme in an unfamiliar accent is perceived as belonging to a different, contrasting category in the listener's native – or for L2 learners, the most familiar – regional accent, thus perceptually shifting the phonetic category the listener hears to a different phoneme than the speaker intended (CS type cross-accent assimilation). For example, CknE pronunciation of /θ/ sounds like an /f/ to an AusE listener. A CG difference between accents would instead mean that the phoneme in an unfamiliar accent is perceived as the same phoneme in the native/familiar accent, but is nonetheless perceived to have a different phonetic quality than that of the native/familiar accent. One possibility is the initial /t/, which has a fricative-like release in CknE. It is unlikely to be perceived as anything but /t/ to AusE listeners but may be recognized as a deviant pronunciation.

Building on the predictions of PAM-L2, these two cross-accent assimilation types, namely, category shifting (CS) and category goodness (CG) differences, were used in the current study. We hypothesized that if L2 learners have acquired English categories, then they would have more difficulty with CS type accent differences than CG type accent differences, relative to the L2 accent they are most familiar with.

2. Experiment 1: AusE versus JaME

2.1 Method

2.1.1 Participants

A total of 16 Chinese native speakers were paid for participating. There were 8 females and 8 males aged from 19;5 to 36;5 (mean age 23;9). They were all university students and had been living in Australia for over 1 year but less than 5 years. All participants reported that they had normal vision and hearing, and that they were familiar with Australian-accented English (AusE). They reported no exposure to Jamaican-accented English (JaME) and Cockney-accented English (CknE). One reaction time outlier had to be removed from the final data set. Thus, the findings reported here included only 15 participants.

2.1.2 Stimuli

All spoken target words were selected from existing corpora of recorded individual words. The recordings were developed for a separate grant-funded project on early development of word recognition across accents. The words in the corpora were organized into high and low frequency words, one and two syllable words, and words that differ between JaME and AusE pronunciations in terms of Category Goodness (CG) or

Category Shifting (CS) differences in one vowel; other phonemes in the words were similarly pronounced in both accents. The CELEX database and the SMH database were used to determine frequency per million of each word. High frequency words were above 40 per million, while low frequency words were below 10 per million. The CS and CG words were selected based on a phonetic mapping table and careful listening at a fine-grained phonetic level to other recordings of the accents available online. The phonetic mapping table was based on a combination of published phonetic descriptions of the accents. We selected to have the type of difference from AusE that we wanted in the target vowel, but minimal differences from AusE in the other segments of each word. Sixty-four monosyllabic words and 64 disyllabic words were selected for the study. Each of the target words was recorded by multiple speakers. There were two female Jamaican Mesolect speakers, and two female Australian speakers producing the same set of words. Each speaker produced the target words multiple times, but only one token of each word from each speaker's recordings was used for the present study. The tokens were selected to be best matched in voice quality and pitch contour between speakers. White noise at an intensity of 35 dB was added to all audio stimuli in the experimental trials (but not the practice trials). This was done to increase the difficulty level of the task.

Visual displays of the response choices on each trial contained four printed words: a target word, an onset competitor, an offset competitor and an unrelated distractor. These were displayed in the four quadrants of the screen and there was also a fifth choice in the centre: "not there". We included the "not there" option to increase the sensitivity of the task (see [2] and [6]). The target word and the onset competitor overlapped in the initial syllable of the word. The offset competitor overlapped with the target only in the final portion of the word, either the rime (monosyllabic words) or the final syllable (disyllabic words). Similar phonemes or orthographic letters never occurred in the same position in the unrelated word as in the target word. In this study, we chose to use printed words rather than pictures for these four response choices. This allowed us to use words that are not easily depictable. A recent study indicated that in this paradigm, printed words elicit similar effects to pictures [10].

All target words were played four times in the experiment, each time by a different speaker. Two of the four occurrences of a word were always in Australian English and the other two occurrences were in Jamaican Mesolect English. Multiple sets of competitor words were used for each target word. This was done to hinder participants from learning competitor sets, and from noticing that the target words appeared more frequently than other words. There was no semantic relationship among any of the words within each set. There were no filler trials in this experiment, but there were 8 practice trials.

2.1.3 Procedure

Participant eye-movements were monitored at a sampling rate of 60 Hz with a Tobii X120 eye-tracker. Participants sat comfortably in front of a computer screen. They placed their chin on a chin rest and their forehead against the top of the frame of the chin rest. The eye-tracker was calibrated to the gaze of each participant. After calibration, participants were shown written instructions on the screen. They were instructed to read the four words shown on the screen (silently) for each trial, and then click on a fixation cross in the center of the screen. As soon as the eye tracker had captured their eyes within the region of interest for the centre fixation, a red square outline appeared around the fixation point, and this triggered the presentation of the audio target stimulus. The audio stimuli were presented over loudspeakers.

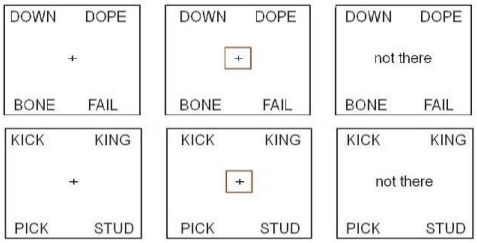

Figure 1. Illustration of the timecourse of the trial procedure for examples of a JaME CS (top) and a JaME CG (bottom) type target word.

All trials were presented in random order. Across trials, the position of the target word, two competitors (onset and offset) and unrelated words were randomized. All target words, competitor words and unrelated words were presented between 14 to 19 times in each quadrant. This was to prevent the participants from focusing on or ignoring a particular quadrant of the screen.

The inter-trial interval (ITI) was set at 500 ms. There were two blocks of trials. Each block contained 256 trials. Participants were put under no time pressure.

2.1.4 Results and discussion

In this study, the proportion of fixation was analysed. The proportion of fixation refers to the proportion of looks to each of the choice words on the computer screen during a given time window [11]. The analysis was limited to the time window from 600 ms to 1600 ms. This is because 600 ms is the time point when the proportion of looks to the centre "not there" dropped below the proportion of looks to the choice words in the corners of the screen; 1600 ms is the time when looking to the target word had reached asymptote.

Figure 2 shows that the proportions of fixation from 600 ms to 1600 ms of target words, onset competitors, offset competitors, unrelated distracters and "not there" in AusE CG, AusE CS, JaME CG and JaME CS. As expected, target words had more fixations across all accents and assimilation types. For the familiar accent (AusE), The AusE target words attracted a greater proportion of fixation than the JaME target words. In both AusE and JaME, the target lines of type CG did not differ much; however, there was a fixation difference for CS accented differences. JaME target words had fewer fixations than AusE target words. The onset competitors of JaME attracted a greater proportion of fixations for CS type accent differences. Take JaME target word DOWN as an example: JaME DOWN would sound like [dəʊn] to listeners. The onset and the nucleus matched with the onset competitor DOPE. Before listeners heard the coda, they were not able to decide which word they were hearing.

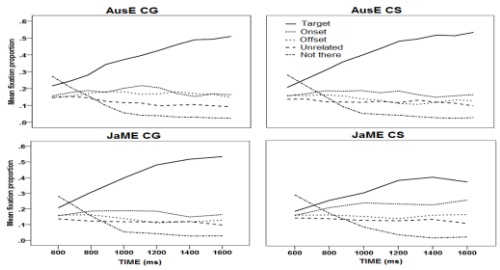

Figure 2. Proportion of fixations over time from 600 ms to 1600 ms to target words, onset competitors, offset competitors, unrelated distractors and "not there" for AusE CG, AusE CS, JaME CG and JaME CS target words.

The mean fixation proportions across the 600-1600 ms window were arcsine transformed for statistical analysis. A three-way repeated measures ANOVA (analysis of variance) with the factors of distractor type (onset competitors, offset competitors or unrelated distractors), accent (AusE Vs JaME) and assimilation type (CG or CS) was conducted on the arcsine transformed values. There were significant main effects of accent [F(1, 14)=15.95, p<.001] and distractor type [F(2, 28)=39.42, p<.001].The main effect of assimilation type was not significant. The accent by distractor interaction was significant [F(2,28)=13.28, p<.001]. The assimilation type by distractor interaction was not significant. The three-way interaction among distractor type, accent, and assimilation type was marginally significant [F(2, 28)=2.71, p=0.084]. To further investigate the three-way interaction separate post hoc ANOVAs were run to evaluate the effect of accent for each distractor type. Onset distractors showed a significant effect of accent [F(1, 14)=39.78, p<.001], but offset and unrelated distractors did not.

Figure 3 compares the mean fixation proportions for each combination of accent and assimilation type in Experiment 1. In general, onset competitors received the highest fixation proportions across all combination of accent and assimilation type. When listeners heard AusE stimuli, they had more fixations, summed across distractor types, for CG targets (onset: M=0.18, SD=0.05; offset: M=0.17, SD=0.05) than CS type targets (onset: M=0.17, SD=0.06; offset: M=0.14, SD=0.04). For the unrelated distractors in AusE, listeners had slightly more fixations for CS differences (M=0.123, SD=0.05) than CG differences (M=0.122, SD=0.04). However, when they heard JaME stimuli, they looked more to all distractor types for the CS targets (onset, M=0.23, SD=0.06; offset: M=0.16, SD=0.05; unrelated: M=0.132, SD=0.04) than they did for CG targets (onset: M=0.19, SD=0.05; offset: M=0.15, SD=0.04; unrelated: M=0.128, SD=0.04).

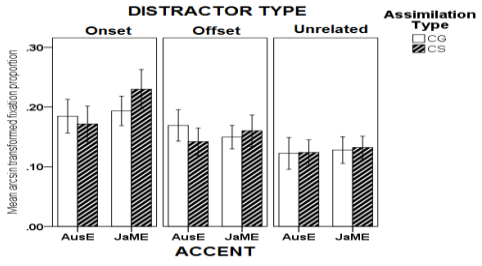

Figure 3. Comparisons of the mean fixation proportions for each combination of accent and assimilation type of onset, offset and unrelated distractors in Experiment 1.

The pattern of assimilation type by accent effects is as follows: When listeners heard JaME stimuli, they showed higher fixations for the CS than the CG words. This result is consistent with the PAM-based prediction that CS words would be more difficult to recognize than CG words for naïve listeners.

3. Experiment 2: AusE versus CknE

3.1.1 Participants

The same group of participants completed the second experiment during the same test session.

3.1.2 Stimuli

The same design of stimulus selection was used for Experiment 2, except that CknE was used as the unfamiliar accent here and the focus was on consonant rather than vowel differences. CknE differs from AusE primarily in consonant pronunciations, so target words were chosen for which one consonant differed in either CG or CS ways in the CknE pronunciation relative to the AusE pronunciation.

3.1.3 Procedure

The same procedure was used as in Experiment 1.

3.1.4 Results and discussion

Figure 4 shows the fixation proportions for AusE CG, AusE CS, CknE CG and CknE CS target words. The target words received a greater proportion of fixations than any distracters across accents and assimilation types. Target word identification was easy in the familiar accent (AusE). AusE CS target words attracted more fixations than CG target words. The same pattern was found in the unfamiliar accent as well. Listeners looked to CknE CS type competitors, particularly when they were in onset position. In the early stage of responses to CknE CS type target words, listeners had more fixations to onset competitors than to target words. For example, a target word such as THRIFT would sound like [frɪft] to listeners. They would easily confuse this with the onset competitor FRILL since the orthography and the phoneme were a better match to the start of the auditory stimulus. At about 800 ms, participants showed an increase in fixations to target words. By this time, they had already heard the entire word and could make the decision based upon the end of auditory the stimulus.

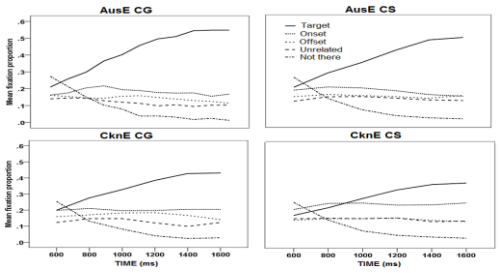

Figure 4. Proportion of fixations over time from 600 ms to 1600 ms to target words, onset competitors, offset competitors, unrelated distractors and "not there" for AusE CG, AusE CS, CknE CG and CknE CS target words.

The fixation proportions over the 600-1600 ms window were again arcsine transformed and a three-way repeated measure ANOVA was conducted. There were significant main

effects of accent [F(1, 14)=24.91, p<.001], assimilation type [F(1, 14)=21.99, p<.001] and distractor type [F(2, 28)=58.70, p<.001]. Meanwhile, the accent by distractor type interaction was significant [F(2, 28)=6.98, p<.01]. The assimilation type by distractor interaction was also significant [F(2, 28)=6.48, p<.01] and distractor by accent by assimilation type interaction was significant as well [F(2, 28)=7.29, p<.01]. Separate post hoc ANOVAs were conducted to evaluate the effect of accent for each distractor type. Again, only onset competitors showed a significant effect of accent [F(1, 14)=20.82, p<.001], but offset competitors and unrelated distractors did not.

Figure 5 shows a comparison of the mean fixation proportions for each combination of accent and assimilation type in Experiment 2. When listeners heard AusE stimuli, they had almost the same amount of fixation to onset competitors for CS (M=0.189, SD=0.04) and CG target words (M=0.188, SD=0.04). They had more fixations to offset and unrelated distracters for the CS (M=0.16, SD=0.04; M=0.14, SD=0.04) than for the CG target words (M=0.15, SD=0.04; M=0.12, SD=0.05). When they heard CknE stimuli, onset competitors and unrelated distracters received more looks for CS (M=0.25, SD=0.05; m=0.15, SD=0.04)) than for CG target words (M=0.20, SD=0.04; m=0.13, SD=0.04).

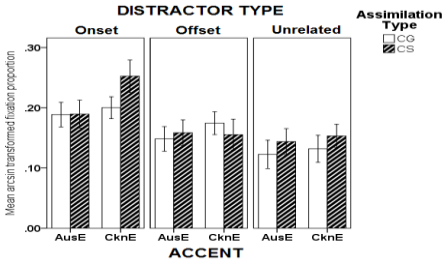

Figure 5. Comparisons of the mean fixation proportions for each combination of accent and assimilation type of onset, offset and unrelated distractors in Experiment 2.

For the CknE targets, CS type received more fixation than CG type for onset competitors. One thing we noticed from our data is that offset competitors received more fixations by our L2 listeners for CG type targets in CknE than in AusE, but there were no CknE-AusE differences for offset competitors for CS type targets, if CG and CS are defined according to the expected AusE listener assimilations of CknE consonants. In addition, a possible contributing factor to why the CG type offset competitors received more fixations than CS type offset competitors in CknE may be the interaction of the white noise added to the stimuli with perception of fricative consonants, as white noise can easily mask fricative spectral differences. More than half of our CG type stimuli started with a fricative consonant. Thus, the listeners may have missed these onsets or may have been uncertain about which fricative they heard, and thus may only have clearly perceived the offsets of these words.

Another important possible contributing factor is that the listeners may have assimilated the CknE and AusE consonants to their L1-Chinese phoneme inventory rather than to L2-AusE phonemes. However, the difference between L1-Chinese and L2-AusE assimilation does not seem very great upon examination. The standard Chinese consonantal inventories include most of the consonants that fall in the CG type of assimilation of the Cockney stimuli by AusE listeners. Most of them could be assimilated similarly to listeners' native Chinese consonant categories. But on the other hand, most of the CS assimilations of CknE consonants by AusE listeners involve consonants that do not exist in Chinese. In AusE-accented words in Experiment 2 (CknE, primarily consonantal differences), the possible AusE to Chinese shifting (CS assimilations) includes [l] to [o], [θ] to [s] and [ð] to [z].

Listeners had equal amounts of fixations to onset competitors for both CS and CG words. This indicates that L2 listeners can set up new phonological categories for their L2, but there are some prerequisites. According to PAM-L2, the establishment of a new L2 category depends on how similar the L1 and L2 phonological categories are. If the two languages share many similarities, then perceptual learning can occur [8]. English consonants do share many similarities with Chinese consonants. Though Chinese does not have the postalveolar affricates /ʧ/ and /ʤ/, it has alveolar and palatal affricates; thus, it should be relative easy for L1-Chinese learners to learn these L2 consonants. Also, our listeners have been immersed in an Australian English-speaking environment for at least a year and up to five years. As their L2 learning experiences increase, they do distinguish at least some of the L2 phoneme contrasts better [6].

4. Summary and conclusions

This study investigated how accent variation, specifically CG and CS type differences between familiar and unfamiliar L2 accents, affect spoken word recognition for Chinese (L2-English) listeners. Our main theoretical prediction, based on PAM-L2 perceptual assimilation principles, was that it would be more difficult for L2 listeners to correctly identify CS than CG target words when spoken in unfamiliar accents, which would be reflected in greater fixations to CS than to CG competitors for JaME and CknE target words. Our study found support for this prediction. This indicates that the L2 listeners have L2 phonological categories that approximate AusE, their familiar regional L2 accent.

We also found that, across accents, onset competitors attracted more fixations than other distracters. The effect of assimilation type was present for both onset and offset distracters for JaME but only for onset competitors in CknE. This difference may be due to the Chinese L2-English listeners relying on their L1 Chinese phonological categories rather than, or in addition to, L2-English categories. Some previous studies [8], [13], [14] and [15] suggest that the activation of false competitors may be caused by the phonemic deviations from the listeners' own accent. In sum, listeners' linguistic experience affects the accurate processing of phonemes in unfamiliar accents.

5. Acknowledgements

This research was supported by research grants from Australian Research Council research grant DP 120104596 (CIs Best & Shaw).

6. References

[1] Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing?. Journal of Experimental Psychology: Human Perception and Performance, 32(5), 1276.

[2] Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probablilistic knowledge about speech production. Journal of Experimental psychology, 35(1), 244-263.

[3] Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. Journal of Memory and Language, 50(1), 1-25.
[4] Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. Journal of Phonetics, 34(2), 269-284.
[5] Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. The Journal of the Acoustical Society of America, 75, 1866.
[6] Brouwer, S., & Bradlow, A. R. (2011). The influence of noise on phonological competition during spoken word recognition. Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China.
[7] Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (1999). Spoken word recognition in the visual world paradigm reflects the structure of the entire lexicon. In Proceedings of the Twenty First Annual Conference of the Cognitive Science Society (pp. 331-336).
[8] Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro & O.-S. Bohn (Eds.) Second Language Speech Learning, pp. 13-34. Amsterdam: John Benjamins Publishing.
[9] Best, C. T. (1995). A direct realist view of cross-language speech perception. Speech perception and linguistic experience: Issues in cross-language research.
[10] Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. Journal of Experimental Psychology: Human Perception and Performance, 37(2), 496.
[11] Tanenhaus, M. K. & Trueswell, J. C. (2011). Eye Movements and Spoken Language Comprehension. In Traxler, M., & Gernsbacher, M. A. Handbook of Psycholinguistics: Elsevier Science. London: Elsevier. pp. 868-869.
[12] Flege, J. E. (1989). Chinese subjects' perception of the word-final English/t/–/d/contrast: Performance before and after training. The Journal of the Acoustical Society of America, 86, 1684.
[13] Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. Journal of Phonetics, 25(2), 169-186.
[14] Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. Journal of Phonetics, 39(4), 479-491.
[15] Weber, A. (2009). The role of linguistic experience in lexical recognition. The Journal of the Acoustical Society of America, 125, 2759.