RAPID SPEECH ADAPTATION AND ITS PERSISTENCE OVER TIME BY NON-STANDARD AND NON-NATIVE LISTENERS

Fengting Du

Newcastle University
f.du4@newcastle.ac.uk

ABSTRACT

Speech produced by L2 talkers is more intelligible to listeners with the same L1 background as the talker than to listeners with a different one. This study extended this line of research to a non-standard native variety and two non-native tonal varieties with similar syllable structures. It examined how these are perceived and adapted to by native and L2 listeners speaking the same L1 or a different L1 from the talker. A post-test after 24 hours detected how the adaptation to accented speech persists. The results found that non-standard and non-native English listeners were better at perceiving and adapting to speech from talkers sharing the same but not a similar or different native background. Such identification accuracy increased after exposure and could continue to improve over a period of sleep. The adaptation effect was not found to extend to new talkers. Additionally, the tested syllabic characteristics were found to have inconsistent effects on speech perception or adaptation.

Keywords: speech perception, speech adaptation, regional- and foreign-accented English

1. INTRODUCTION

Many studies have investigated the influence of both linguistic and non-linguistic traces of a foreign accent on speech perception. The speech identification accuracy or the intelligibility of accented speech to native speakers relates to phonetic and phonological features, such as segmental properties and prosodic patterns [11, 12, 14, 15, 16], and demographic characteristics like race and social status [13, 17]. Recently, more literature has investigated whether an intelligibility difference exists in cross-linguistic communication [2, 3, 5, 7, 8]. These studies found that when an L2 talker talks to an L2 listener from the same native language background, it will give the listener an intelligibility benefit.

Listeners familiar with foreign accents or language teaching experience were found to view L2 speech more favourably. They are likely to exert less effort in accented speech processing and be able to understand more speech that is less intelligible to others without accent familiarity. Additionally, it has been found that exposure to systematic variability of foreign accents facilitates the generalisation of accent features, which can develop into the adaptation to foreign-accented speech [2, 8]. Studies have shown that listeners can adapt to both talker-specific and talker-varied accented speech, as well as accent-independent adaptation. Moreover, studies also found that greater talker variabilities in the exposure can lead to better adaptation.

Although the foreign-accented speech is usually labelled as hard to understand, it has been found that listeners can adapt to accented speech with 2-4 sentences, about one minute of exposure to the target accent [3]. Studies also found that listeners' perceptual ability can be 'improved' if they put in more effort, further suggesting that the intelligibility of accent speech depends not only on the talker but also on the listener [9, 10]. Such inconsistency also shows that speech perception is a complex progress involving a range of interlocuters' language-related background factors, such as native language, L2 proficiency level, language experience, language attitudes and so on.

Many studies compared the intelligibility of native speech with that of L2 speech produced by the talker having the same native language background as the listener. However, such studies cannot answer where the speech intelligibility benefit comes from. Is it from the same articulatory features of segments or the same phonological change transferred into L2 production? To answer these questions, this study examines the scenarios where the interlocutors share a similar L1 background (e.g., having similar syllable structures in terms of the consonant clusters in the onset and coda positions) and where the talker is a native speaker having a regional accent. Additionally, the persistence of rapid speech adaptation is understudied but interesting to look at as it can shed some light on the mechanisms of maintenance with short-term memory.

Therefore, the present study aims to investigate how regional and foreign English varieties are perceived and adapted to by native listeners, L2 listeners with the same L1 background as the talker and L2 listeners with a different L1 background from the talker. It also touches on how the adaptation will change over time and investigates the role of language proficiency in perceiving and adapting to

regional and foreign-accented speech. Burmese and Mandarin accented learners were enrolled, as both their L1s are monosyllabic and only permit consonant clusters in the syllable onset but not the coda [1, 4]. The native variety studied is Geordie, a dialect spoken by people living in Newcastle, England.

2. MAIN EXPERIMENT

2.1. Participants

Six female participants were recorded. They were talkers with the three tested accent backgrounds under study (two from each).

Twenty-one listeners (five Geordie, five Mandarin advanced and five Mandarin intermediate and six Burmese intermediate English learners) aged 19-25 were enrolled and remunerated £10 for participation. The English proficiency of L2 learners was categorised as intermediate if their latest score of *The International English Language Testing System* (IELTS), a standard test to measure L2 learners' English proficiency level, was between 6-7 out of 9 and advanced if it was above 7.

2.2. Materials

2.2.1. Lexical transcription

Twelve lists of 20 words were selected from the reading tests of IELTS. In a lexical list, ten of them had a consonant cluster at the onset position, while the rest ten with a cluster at the coda. Most stimulus words were monosyllabic or disyllabic, with only a few trisyllabic.

2.2.2. Accent exposure

Six picture pairs with 12 differences between each pair were taken from *London UCL Clear Speech in Interaction* (LUCID Corpus) [6] and used as exposure materials. Participants saw a picture and heard someone (with one of the accents under study) describing how the picture differs from another in a spot-the-difference task. In the picture presented, 20 parts were numbered randomly, 12 target parts and eight distractors. To ensure consistency in the linguistic content and exposure time, the researcher wrote the scripts for the recordings. Each recording lasted about 1.5 to 2 minutes long.

2.3. Procedures

This online experiment was built on *LabVanced* and had seven phases in two days. The participant was required to find a quiet place with their computer connected to the Internet and wear earphones if necessary. The first-day experiment had six tests: four main tests involving lexical transcription (pre-test: P0 and three post-tests: P1, P2, P3) and two exposure tests (before P1 and P2, respectively).

Each main task had three 20-word lists recorded in each accent under study. In each word list, half words had a consonant cluster in the onset, and the other half had a cluster in the coda. Stimulus words in P1 and P2 were all recorded by the same talker with each accent as in P0. Listeners were exposed to a new talker of each accent in P3 to see if listeners could generalise the idiosyncrasies of the studied accents to achieve talker-independent perception. One point was awarded if the participant could correctly hear and transcribe a word.

Exposure tasks aimed to help the participant focus on speech decoding. The accuracy from such tasks was only used as criteria to remove the outlier data but not in data modelling.

Another post-test was conducted 24 hours away from the first exposure. The participant transcribed a combined word list of 40 tokens produced by the talker from exposure and the novel one (coded as P4 and P5, respectively). It was scored in the same way as the other post-tests.

2.4. Analysis

The data from one Burmese participant was removed as an outlier. The variables expected to affect the accuracy of perceiving accented speech were the L1 relationship between the interlocutors (same, similar or different), the position of the consonant cluster (onset or coda), test phases (P0, P1, P2, P3, P4 and P5), the talker (same as or different from the exposure recording) and the language proficiency of the listener (intermediate-I, advanced-A). To be specific about the relationship between two L1s, a listener listening to a talker from the very native language background was coded as the same. An L2 listener listening to an L2 talker was coded to have a similar language background. Other scenarios were categorised as having a different background between interlocutors (Geordie listeners listening to Burmese or Mandarin talkers and vice versa).

The accuracy was analysed under separate conditions: pre-test, adaptation to and generalisation of accented speech and the role of language proficiency. Although each condition would not involve all variables, they were all modelled using mixed-effect models. The inclusion of variables was based on theoretical relevance and the significance of the added effects achieved by the likelihood ratio test. The basic model had two random effects: the participants and lexical stimuli. All modelling started from a null model and added one relevant fixed effect each time. The variable stayed in the model only

when it contributed significantly to model fit, the same as the interactions among variables.

3. RESULTS

3.1. Performance across the whole experiment

Fig. 1 illustrates the performance of each listener group across the tests. Generally, Burmese-accented speech achieved the lowest perceptive accuracy. The perceptive accuracy of Geordie and Burmese accents kept increasing from P0 to P2, suggesting a potential adaptative effect. Although the accuracy of Mandarin-accented speech seems to decrease, it may be due to fatigue, as it was the last test in each testing round. The Geordie listeners had the highest accuracy in perceiving three accents and nearly perfect in perceiving their own accented speech. The perceptive performance of the Mandarin-A listeners was lower than that of the Geordie listeners but higher than the Burmese and Mandarin-I listeners. The perception of a different talker's speech was lower than the same.

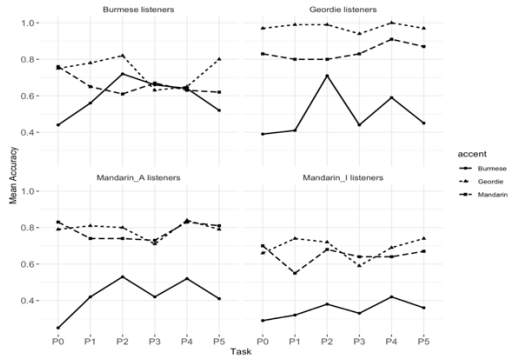

Figure 1: Performance of listeners in each scenario.

3.2. Pre-test

The relevant fixed effects to model pre-test performance were the L1 relationship and the consonant cluster position. The addition of the language background of interlocutors significantly contributed to data likelihood ($\chi^2(5) = 19.5$, $p < .0001$), but the addition of the consonant cluster position did not ($\chi^2(6) = 0.18$, $p > .1$).

Table 1 summarises the model's output. It shows that compared with the listener with a different native language background from the talker, the listener with the same L1 performed better in perceiving speech (running a Plogis function generated an 80.7% chance of transcribing correctly). By comparison, the listener with a different L1 had a lower performance (66.1%), slightly higher than the one with a similar L1 (65.3%).

Table 1: Output from the pre-test model.

Fixed effect	Estimate	SE	z-value
Intercept	0.67	0.31	2.15
Same L1	0.77	0.22	3.47
Similar L1	-0.03	0.26	-0.13

3.3. The adaptation to and generalisation of accented speech

In modelling the adaptation to accented speech, the potential fixed effects were added in the order of the L1 relationship, the test phase and the consonant cluster position. The test phases included were only P0, P1, P2 and P4, as the recordings were from the same talker. The final model included the L1 relationship ($\chi^2(5) = 110.27$, $p < .001$), the test phase ($\chi^2(8) = 13.70$, $p < .01$) and the consonant cluster position ($\chi^2(9) = 5.02$, $p < .05$), without detected interactions. Table 2 summarises the model output.

Table 2: Output from the adaptation model.

Fixed effect	Estimate	SE	z-value
Intercept	0.15	0.34	0.44
Same L1	1.06	0.14	7.61
Similar L1	-0.16	0.16	-0.97
Onset cluster	0.64	0.28	2.28
P1	0.61	0.37	1.66
P2	0.94	0.38	2.46
P4	1.28	0.36	3.51

The outputs demonstrate that compared with the interlocutors having different L1 backgrounds, the intelligibility of accented speech produced by the talker with the same L1 background was higher. However, a similar native language background did not benefit the listener's perception, if not lowered it. It was contrary to the hypothesis that L2 learners may share similar learning strategies in language learning, making L2 learners familiar with other L2-accented speech features. The results also suggest that it was easier for the listeners to transcribe words that had an onset consonant cluster.

The effect of the test phase significantly influenced the listener's perceptive performance in that the speech perception accuracy increased over the experimental progress. A higher probability of correctly perceiving accented speech in the first two post-tests demonstrates the instant benefits of short exposure to phonological processing. In the post-tests starting 24 hours after the start of the experiment, the accentual processing ability of the listeners kept improving. It indicated that the listeners could extract and generalise the accentual characteristics and retain

their perceptual adaptation to improve the efficacy of decoding accented speech in the short future.

To model if the listener could generalise talker-variability from an individual talker to match a new talker's speech characteristics in phonological processing, the L1 relationship, the test phase, the talker and the cluster position were involved. The test phases involved were P2, P3, P4 and P5, as the exposure factor was controlled. The final model included the L1 relationship ($\chi^2(5) = 116.34$, $p < .0001$) and the talker ($\chi^2(6) = 18.97$, $p < .0001$).

Table 3: Output from the adaptation model.

Fixed effect	Estimate	SE	z-value
Intercept	1.10	0.22	5.04
Same L1	0.91	0.13	7.21
Similar L1	-0.13	0.14	-0.90
Different talker	-0.45	0.10	-4.39

The significant talker factor indicates that brief exposure to accented speech from a single talker may not be sufficient for listeners to develop a talker-independent accentual adaptation.

3.4. The role of language proficiency of the listener

In modelling the role of language proficiency, only the data from two groups of Mandarin listeners were analysed. The potential fixed effects were the same as those modelled in the last section analysis except for the inclusion of the proficiency effect. The final model of the adaptation effect contained the fixed effects of the L1 relationship ($\chi^2(5) = 83.94$, $p < .001$) and the proficiency $\chi^2(6) = 5.64$, $p < .05$).

Table 4: Output from the L2 proficiency model.

Fixed effect	Estimate	SE	z-value
Intercept	2.15	0.30	7.18
Same L1	-0.31	0.31	-0.99
Similar L1	-2.60	0.31	-8.39
Proficiency-I	-0.79	0.29	-2.76

As shown in Table 4, the intermediate Mandarin listeners performed worse than their advanced peers, which is expected given the known advantage of higher proficiency in L2 learners. However, the effect of the L1 relationship in this model differs from the previous results. When Mandarin talkers and listeners shared the same L1, it appeared to have a negative effect on their perceptual accuracy, as with other interlocutors who shared similar L1s. The factor of cluster position was found to be insignificant, suggesting that it does not consistently influence accentual perception. The test phase was also insignificant, indicating that taking a one-day break did not result in significant perceptual improvement.

In Table 5, the generalisation effect model is shown. It is similar to the previous results but includes the proficiency level variable. The results indicate that advanced listeners had better perceptual accuracy ($\chi^2(7) = 6.01$ and $p < .05$). Mandarin listeners were unable to transfer accentual adaptation to a novel talker after minimal accentual exposure as well.

Table 5: Output from the adaptation model.

Fixed effect	Estimate	SE	z-value
Intercept	1.44	0.30	4.77
Same L1	-0.11	0.34	-0.33
Similar L1	-2.15	0.34	-6.40
Different talker	-0.55	0.14	-4.00
Proficiency-A	0.73	0.25	2.88

4. DISCUSSION AND CONCLUSION

The results of the current study demonstrate that the listener can adapt to accented speech quickly. It also sheds further light on how the relationship between the native languages of the interlocutors influences phonological processing. One limitation is that it only tested the role of listeners' proficiency and did not strictly control the L2 talkers' proficiency, which may cause the generally low intelligibility of Burmese-accented speech in the study. Although the research included native talkers whose English proficiency levels were naturally higher than L2 learners', the proficiency of L2 learners still should have been controlled to avoid the perceptive variation caused by different L2 intelligibility levels.

The significant talker effect on perceptual accuracy implies that listeners may not be able to transfer accentual adaptation to a novel talker through minimal accentual exposure. It is crucial to note that the talker effect may also arise from the surprise effect of a change in voice. Therefore, further investigation is needed in order to determine what will occur once the surprise effect wears off.

The inconsistency regarding the cluster position factor between analysing Geordie, Burmese, and Mandarin advanced groups and only two Mandarin groups implies that the impact of cluster position on accented speech perception could vary depending on individual factors, such as L1 background and proficiency level. Thus, more work should be undertaken to investigate the perception of and adaptation to L2 accented speech in more contexts and the roles of more aspects of previous linguistic knowledge in driving accentual adaptation, such as phonetic property and prosodic patterns.

5. ACKNOWLEDGEMENTS

I thank my supervisors, Dr. Rory Turnbull, Dr. Niamh Kelly and Prof. Martha Young-Scholten, for their guidance and support. I am also grateful to all informants and participants involved in this research and members of the Phonetics & Phonology Research Group at Newcastle University who provided constructive feedback on an earlier draft.

6. REFERENCES

[1] Watkins, J. W. 2001. Burmese. *Journal of the International Phonetic Association* 31(2), 291–295.

[2] Bent, T., A. R. Bradlow. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114(3), 1600-1610.

[3] Clarke, C. M., Garrett, M. F. 2004. Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America* 116(6), 3647-3658.

[4] Duanmu, S. 2007. *The Phonology of Standard Chinese.* Oxford University Press.

[5] Bradlow, A. R., Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition* 106(2), 707-729.

[6] Baker, R., V. Hazan. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43(3), 761-770.

[7] Grynpas, J., Baker, R., Hazan, V. 2011. Clear speech strategies and speech perception in adverse listening conditions. *International Phonetic Association* XVII, 17-21.

[8] Baese-Berk, M. M., Bradlow, A. R., Wright, B. A. 2013. Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America* 133(3), EL174-EL180.

[9] Pichora-Fuller, M. K., Kramer, S. E., Eckart, M. A., Edwards, B., Hornsby, B., Humes, L. E.,...Wingfield, A. 2016. Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing* 37, 5S–27S.

[10] Peng, Z. E., Wang, L. M. 2019. Listening effort by native and non-native listeners due to noise, reverberation, and talker foreign accent during English speech perception. *Journal of Speech, Language, and Hearing Research* 62(4), 1068-1081.

[11] Steffman, J., Jun, S. A. 2019. Perceptual integration of pitch and duration: Prosodic and psychoacoustic influences in speech perception. *The Journal of the Acoustical Society of America* 146(3), EL251-EL257.

[12] Alves, U. K., Brisolara, L. B. 2020. Listening to accented speech in Brazilian Portuguese: On the role of fricative voicing and vowel duration in the identification of/s/–/z/minimal pairs produced by speakers of L1 Spanish. *Journal of Portuguese Linguistics* 6, 1-23.

[13] Kutlu, E. 2020. Now you see me, now you mishear me: Raciolinguistic accounts of speech perception in different English varieties. *Journal of Multilingual and Multicultural Development*, 1-15.

[14] Baese-Berk, M. M., Drake, S., Foster, K., Lee, D. Y., Staggs, C., Wright, J. M. 2021. Lexical diversity, lexical sophistication, and predictability for speech in multiple listening conditions. *Frontiers in Psychology* 12, 661415.

[15] Severijnen, G. G., Bosker, H. R., Piai, V., McQueen, J. M. 2021. Listeners track talker-specific prosody to deal with talker-variability. *Brain Research* 1769, 147605.

[16] Steffman, J., Katsuda, H. 2021. Intonational structure influences perception of contrastive vowel length: The case of phrase-final lengthening in Tokyo Japanese. *Language and Speech* 64(4), 839-858.

[17] Kutlu, E., Tiv, M., Wulff, S., Titone, D. 2022. The impact of race on speech perception and accentedness judgements in racially diverse and non-diverse groups. *Applied Linguistics* 43(5), 867-890.