Second-language experience and speech-in-noise recognition: the role of L2 experiencein the talker-listener accent interaction.

Melanie Pinet [1], Paul Iverson [1] and Mark Huckvale [1]

[1] Department of Speech, Hearing and Phonetic Sciences, University College London, UK

m.pinet@ucl.ac.uk, p.iverson@ucl.ac.uk, m.huckvale@ucl.ac.uk

Abstract

This study investigated how L2 experience modulates L1-L2 talker-listener intelligibility. L1 southern British English (SE) and L1 French listeners with varying L2 experience (Inexperienced 'FI', Experienced 'FE' and Bilinguals) were tested on their speech-in-noise recognition of English sentences that were spoken with a range of accents (SE, FI, FE, Northern Irish and Korean-accented English). Results showed that while the FI listeners had graded sensitivity for the accents, the SE listeners' recognition processes were selectively tuned to their own accent. Overall, this suggests that L2 experience affects talker-listener accent interactions, altering both accent intelligibility and selectivity of accent processing.

Index Terms: L2 speech perception, L2 experience, talker-listener accent interaction, L2 proficiency, acoustic similarity.

1. Introduction

Speech recognition in noise critically depends on an interaction between the accents of the talker and the listener. Listeners can be accurate at recognizing a wide range of accents in quiet, but in noise they are much poorer when they try to understand native (L1) or non-native (L2) accented speech that does not closely match their own accent. L1 listeners have a marked advantage when listening to speech produced by other L1 talkers than by L2 talkers [1, 2], especially when they share the same L1 accent [3], but L2 listeners can actually be better at recognizing L2 than L1 speech, particularly when the L2 talkers and listeners share the same L1 [4, 5]. This interaction is well established, but it is still unclear why it occurs.

One possibility is that the interaction could be due to the familiarity or experience a listener has with the talker's accent. For instance, Adank *et al*. [3] found that the processing cost associated with listening to an unfamiliar L1 accent was greater compared to a familiar L1 accent. In their study, Standard Southern British English (SE) listeners could adapt faster to SE-accented speech than Glaswegian English (GE), but GE listeners could process SE speech as fast as their own accent; as it is the dominant accent in the UK and therefore GE listeners are familiar with it.

Likewise, several studies have shown that L1 listeners are able to readily adapt to L2 accents even though the initial processing speed is slower compared to L1 speech [6]. This rapid adaptation is comparable to an effect of familiarity because the listeners become more familiar with the accent as the study progresses.

This familiarity effect can also be related long term L2 experience. Recent research [5, 7] has shown that L2 listeners who are highly experienced with L1 speech are more accurate at recognizing L1 accents, while more inexperienced L2 listeners have more of an advantage for L2-accented speech, being better with similarly accented L2 talkers.

Another possibility is that L2 accents are more intelligible to L2 listeners because there are listening strategies available to them that mutually increase the intelligibility of their speech. Bent and Bradlow [4] argued that L2 talkers are more intelligible to L2 listeners when they share an interlanguage, i.e. when they share the same L1. This is because they are more 'equipped' to interpret the acoustic-phonetic features in the speech of the L2 talkers even though they may deviate from the target language because they share the phonetic and phonological knowledge of both their L1 and L2. However, it is also plausible that L2 speech is more intelligible to L2 listeners because they benefit from more global listening strategies. For instance, in Bent and Bradlow's study [4], L1 speakers of Chinese who had learned English as an L2 were more accurate at recognizing both Chinese- and Korean-accented English than L1 English speech in noise. Thus, the intelligibility benefit for L2 speech could be due to L1 features shared between the talker and the listener or a listening strategy more broadly available to L2 speakers.

Alternatively, it is also very probable that the interaction could be driven by the basic acoustic similarities in the accents of the talker and the listeners. Studies in which listeners shared L2 accent acoustic-phonetic features [4, 5] showed enhanced mutual intelligibility, while other studies [8] didn't replicate such findings of L2 intelligibility benefit because of a potential lack of shared acoustic similarities between the two L2 accents. Likewise, some of the accent familiarity effect due to L2 experience could also be driven by acoustic similarity between the L2 listener and the L1 talker [5]. Indeed, as L2 listeners become more proficient in their L2, their own production becomes closer to L1 speech, and therefore the acoustic similarities shared between the L1 talker's and the experienced L2 listeners' accent could enhance intelligibility.

The aim of the present study was to investigate how L2 experience affects the talker-listener accent interaction in speech in noise recognition. L1 French listeners with varying L2 English experience (inexperienced, FI; experienced, FE; and bilinguals, FB) and L1 southern British English (SE) were tested on the recognition of English sentences mixed in speech-shaped noise that were spoken with a range of L1 and L2 accents (SE, FE, FI, Northern Irish-accented English and Korean-accented English). L2 talkers and listeners with matching and mismatching L1s and L2 experience were tested in order to observe any effect of intelligibility benefit due to a shared interlanguage or advantage for L2 speech and effects of L2 experience. L1 and L2 listeners were tested on a variety of L1 and L2 accents that matched and mismatched their own in order to observe any effects of perceptual facilitation due to accent familiarity and/or interlanguage effects.

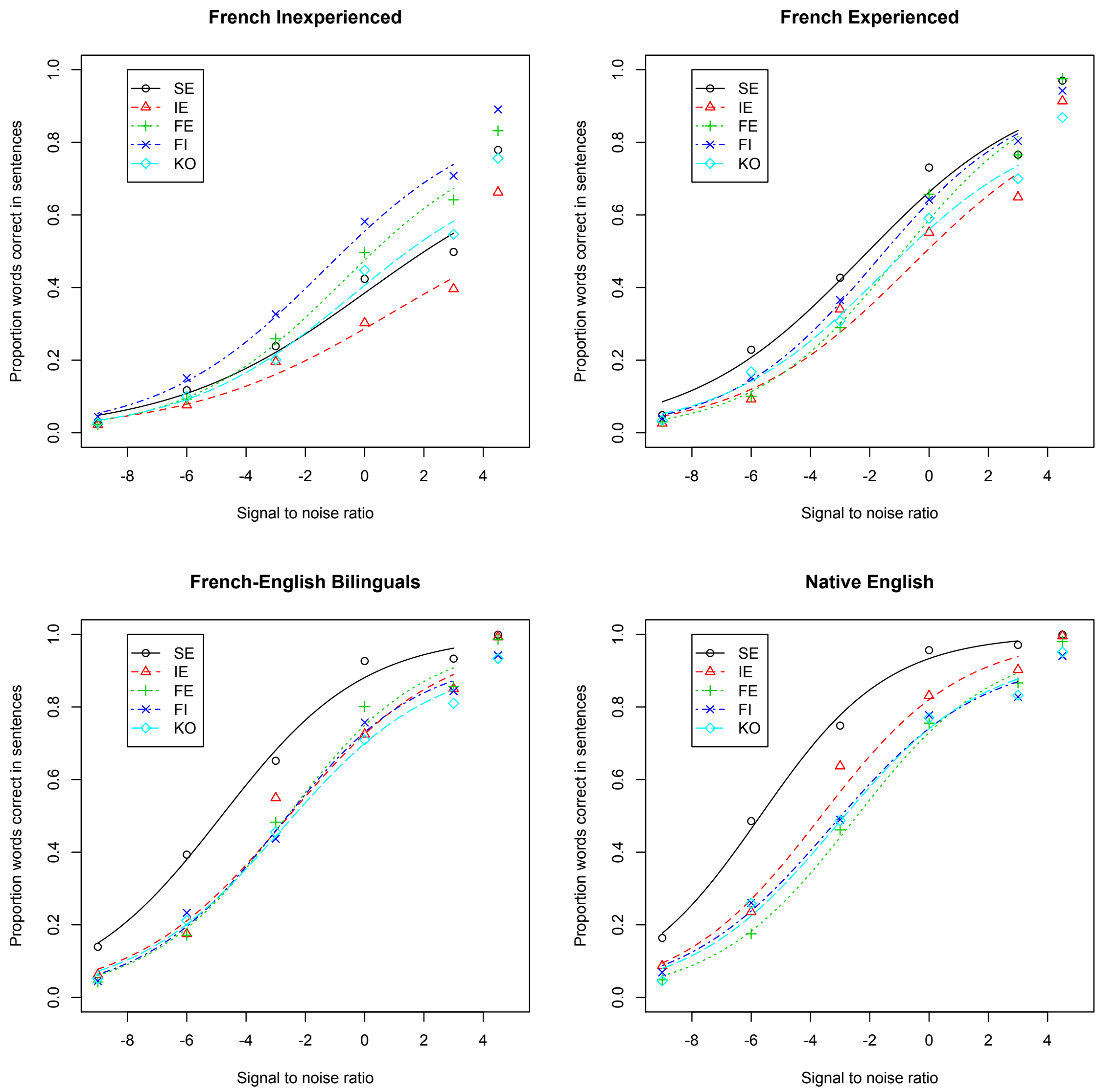

Fig. 1. Psychometric functions of identification performance for the FI, FE, FB and SE listeners across noise levels. Single points on the right hand side represent the listeners' performance in quiet.

2. Method

2.1. Subjects

2.1.1. Listeners

A total of 93 subjects took part in the study: a group of 21 monolingual native Southern British English listeners (SE) aged 18 to 48 years (mean: 28 years) were tested in London. Three group of French native listeners with differing experiences with English were also tested: a group of 16 French-English bilingual listeners (FB) aged 18-36 years (mean: 21 years) were tested in London. They had acquired both English and French from birth or at a very young age (age range of acquisition of French: 0 to 18 months, mean: 2.6 months; age range of acquisition of English: 0 to 9 years, mean: 11 months) and had a native-like command of both languages (their spoken fluency was assessed by the experimenter who is a native French speaker). A group of 24 French Experienced (FE) listeners, aged 18 to 48 years (mean: 25 years), were also tested in London. They had been residing in an English speaking country for a period of time ranging from 1 month to 8 years (mean: 15.5 months) and had started learning English at school from the age of 6 to 14 years old (mean: 11 years). Both groups of listeners were living in London at the time of testing and were experienced with Southern British English. Finally, a group of 32 French Inexperienced (FI) listeners, aged 18 to 54 (mean: 25), were tested in Northeastern France. They had started studying English at school from the age of 7 to 13 years old (mean: 11 years); the subjects had spent little time in English speaking countries (i.e., no more than 8 weeks).

2.1.2. Talkers

Four talkers (two males and two females) of Standard Southern British English (SE), Northern Irish English (IE), Korean-accented English (KO), French Experienced-accented English (FE) and French Inexperienced-accented English (FI) were recorded reading the Bamford-Kowal-Bench (BKB) sentences [9]. The FE and FI talkers matched the French listeners in terms of L2 experience and spoken proficiency. The digitized recordings were embedded in speech-shaped noise with -9, -6, -3, 0 and +3 dB signal-to-noise (SNR) ratios. The speech-shaped noise was generated for each individual talker by first concanating all of the talker's recordings, then calculating and smoothing the spectrum across all the sentences and filtering the generated noise to obtain the same spectral shape of the material.

2.2. Procedure

The subjects performed a sentence identification task where they listened to the stimuli and repeated what they had heard. All stimuli were played to the subjects using a laptop over headphones. Responses for tests were given verbally (i.e., the experimenter marked how many keywords were spoken correctly). Each sentence was presented only once (i.e., they were not repeated within or across conditions). To make speech recognition challenging and to avoid any effects of accent tuning, the subjects heard the stimuli in random accent and SNR (including quiet) order. The subjects were given a practice session of 16 stimuli at the start of the experiment to familiarise themselves with the test, then two blocks of 140 stimuli with 56 sentences for each of the five accent conditions mixed within each block. The order and sentence assignment to conditions were counterbalanced between subjects.

3. Results

Figure 1 displays the listeners' recognition accuracy (i.e., the proportion of words correctly identified in the sentences) for the five accent conditions across all noise levels. The graphs show psychometric functions fitted to the data from 0 to the top scores from the average percentage correct scores for the stimuli presented in quiet and across the different noise levels (i.e., -9, -6, -3, 0 and +3 dB SNR). The single symbols represent the listeners' performance in on the stimuli presented in quiet.

Overall, Figure 1 shows that the listeners responded to the various accents in different manners and their overall performance on the task was clearly affected by their experience with English; the SE listeners performed the best, followed by the FB, then FE and the FI listeners performed worst overall. In order to test the differences of performance on the various accents between the groups, a repeated-measures ANOVA was conducted with accent condition as within-subject factors and listener group as a between-subjects factor. All analyses were carried out on arcsine-transformed scores. For the stimuli presented in noise, the percentage of correct answers was averaged across all noise levels in each accent condition to obtain each subject's performance threshold. Scores in quiet were calculated separately so that analyses for the quiet and noisy listening conditions were carried out separately. Only the results for the stimuli presented in noise will be presented here. The results revealed significant main effects of listener group, F (3, 89) = 86.05, $p < .01$, accent condition, F (4, 356) = 89.44, $p < .01$, and a significant interaction between accent condition and listener group, F (12, 356) = 41.53, $p < .01$. This confirmed that the group of SE listeners performed better overall and as the L2 listeners' experience with English decreased, so did their recognition accuracy.

In order to further investigate the interaction between the accents of the talkers and listeners and assess the effects of the accent stimulus material on the relative intelligibility of the talkers' accents on the listeners, ANOVAs and Tukey tests were conducted separately for each group of listeners. First, Figure 1 shows that the FI listeners displayed graded levels of recognition accuracy. They were clearly affected by all of the accents, being most accurate at recognizing sentences produced by similarly strongly accented FI speakers (50% recognition accuracy at 1dB SNR), then FE speakers, and intelligibility became gradually poorer for the other accents. This was confirmed by a significant main effect of accent condition, F (4, 124) = 61.40, $p < .01$. Tukey tests were run to reveal any significant differences in performances between accent conditions. The tests revealed that listeners found all accents to be significantly different in terms of intelligibility, $p < .05$, except SE and KO speech.

For the FE listeners, Figure 1 shows that they had less of a graded sensitivity to the accents and performed better on SE speech (50% recognition accuracy at 2dB SNR), then FI speech, and were least accurate at recognizing IE speech (50% recognition accuracy at 0dB SNR). The analysis revealed significant main effects of accent in noise, F (4, 92) = 26,57, $p < .01$. In addition, Tukey tests revealed that all accents were significantly different from each other, $p < .05$, except KO and IE speech and KO and FE. Therefore, even though the FE listeners displayed less graded sensitivity to the accents and the differences in accent intelligibility were reduced compared to the FI listeners, it is not the case that they were all recognized the same.

For the SE listeners, figure 1 shows that they were selectively tuned to their own accent and were most accurate at recognizing SE speech (50% recognition accuracy at ~6dB SNR) and IE speech was only marginally more intelligible than the other L2 accents (50% recognition accuracy at ~3.5dB SNR versus ~3dB SNR), and the analysis revealed a significant main effect of accent, F (4, 80) = 89,78, $p < .01$. Tukey tests showed that the listeners performed significantly better on SE speech and performed equally poorly on the L2 accents. Indeed, all accents were significantly different from each other, $p < .05$, besides FI and FE speech, KO and FE speech and KO and FI speech. Interestingly, IE speech was significantly less intelligible than SE speech in noise, $p < .05$, even though both accents were equally intelligible in quiet, and marginally more intelligible than the L2 accents. Therefore, adding noise to the stimuli made the SE listeners' recognition processes selectively tuned to their own accent and perform at a similar level on the other accents.

Finally, for the FB listeners, Figure 1 displays a similar pattern of recognition to the SE listeners'; they were most accurate at recognizing sentences produced by SE-accented speakers (50% recognition accuracy at ~5dB SNR) but displayed no sensitivity to the other accents (performing at 50% recognition accuracy at ~2.5dB SNR for all of them). This was confirmed in the analysis by a main effect of accent, F (4, 92) = 43,58, $p < .01$. Tukey test revealed that only SE speech was significantly different from all the other accents, $p < .05$. All the other accents were equally intelligible: FI and FE speech, IE and FE speech, IE and FE speech, KO and FI speech, and KO and IE speech. Therefore the FB listeners were selectively tuned to SE speech and show no differences for the other accents. Interestingly, adding noise to the stimuli significantly decreased their performance on accents in which they were performing at ceiling level in quiet.

4. Discussion

The results demonstrated a clear talker-listener interaction; the SE listeners were more accurate at recognizing speech produced by the SE talkers but showed no differences for the L2 accents and were only marginally better on IE speech. They were thus selectively tuned to their own accent. The FI listeners, on the other hand, performed better on FI-accented speech and became progressively worse on the other accents, thus showing some graded sensitivity. The more experienced L2 listeners (FE and FB) were better at SE speech in noise, becoming selectively tuned to it as their experienced with L1 speech increased. Therefore, the results showed that the interaction was strongly modulated by the listeners' L2 experience.

It is possible to consider that accent familiarity does play a role in the data because the listeners more experienced with SE speech performed the best on this accent. However, accent familiarity doesn't account for how the listeners performed on the other accents. For instance, it can't explain for the experienced L2 listeners' poor recognition accuracy of French-accented speech in noise, which is surprising given that all the listeners had a high familiarity with this accent. Indeed, the FB listeners all reported having some familiarity with French-accented speech having been raised in a mixed French and English speaking environment and having a French-accented parent or family member and French-accented peers through the community they live in (i.e., FE speakers living in the UK or FI speakers living in France). Yet, the noisy listening conditions made their recognition of French-accented speech drop drastically despite their familiarity with this accent and native-like command of French, and instead they were selectively tuned to SE speech. Likewise, accent familiarity couldn't account for the fact that the FI listeners performed equally well on SE and KO-accented speech since there is no particular reason why they would have had as much exposure to KO-accented speech as SE speech. In fact, it is quite unlikely that they had any exposure to KO speech at all, but instead had some exposure to SE speech through the media and short travels to the UK.

One could also imagine that the results show an interlanguage benefit [4] because the FI listeners displayed a clear advantage for FI speech. However, there was no evidence that L2 listeners had a particular advantage for L2 speech. Indeed, similarly as in Stibbard and Lee [8], the interlanguage benefit didn't occur when the L1s of the L2 talkers and listeners were too far apart: none of the L2 listeners showed a benefit for the KO talkers' speech because their accents had a lack of shared phonetic similarities. In addition, the results showed the interlanguage benefit is very dependable on L2 experience. Indeed, the FE and FB listeners did not show a clear advantage for French-accented speech over SE-speech despite having an intact French phonological system and being L1 French speakers or bilingual from birth in French. Consequently, it is surprising that they didn't perform as well on the French accents as on SE speech throughout all listening conditions, and likewise, that they were unable to recruit their knowledge of the French phonological system to overcome the challenging listening conditions. Instead, the FB listeners were selectively tuned to SE speech, whilst the FE listeners started to display some selectivity for it as well.

The FB listeners' results described above were indeed surprising given their intact French phonology and experience with French accented speech, and this could be due to the blocking design. Indeed, the stimuli were presented in a mixed-accent listening task in order to avoid any accent tuning effects in the first instance, but it is possible that, due to the blocking design, the more experienced listeners could only listen to one accent. However, previous research has shown that listeners can rapidly adapt to a novel or L2 accent [4] therefore the listeners should have been able to better adapt to the FI accent, but in this listening context, they were only able to be selectively tuned to SE speech. Future studies will further examine the issue of accent adaptation by observing differences of testing listeners' accent perception in a mixed-versus one-accent blocking design to observe effects of adaptation.

5. Conclusions

To conclude, this study investigated how the talker-listener interaction is modulated by L2 experience. Overall, the results suggest that L2 experience affects strongly the talker-listener accent interactions, altering both the intelligibility of different accents and the selectivity of accent processing. It demonstrated that as L2 experience increases, L2 listeners start to lose any graded sensitivity to accents they may have had and begin to specialize for L1-accented speech. It also showed that L1 and bilinguals from birth listeners are highly specialized for L1 speech and selectively tuned to it. Finally, preliminary data indicated that there were strong correlations between speech-in-noise recognition and the acoustic similarity of the talkers' and listeners' accents.

6. Acknowledgements

The project research was supported by the MRC and ESRC research councils.

7. References

[1] Munro, M. (1998). "The effects of noise on the intelligibility of foreign-accented speech," Studies in Second Language Acquisition 20, 139-154.

[2] van Wijngaarden, S. J. (2001). "Intelligibility of native and non-native Dutch speech," Speech Communication 35, 103-113.

[3] Adank, P., Evans, B., G., Stuart-Smith, J. and Scott, S., K. (2009). "Comprehension of familiar and unfamiliar native accents under adverse listening conditions", Journal of Experimental Psychology: Human Perception and Performance 35, 520-529.

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

[5] van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (2002). "Quantifying the intelligibility of speech in noise for non-native listeners," Journal of the Acoustical Society of America 111, 1906-1916.

[6] Clarke, C. M. and Garrett, M. F. (2004). "Rapid adaptation to foreign-accented English", Clopper, C.G., 116, 3647-3658.

[7] Pinet, M. and Iverson, P. (under revision) "Talker-listener accent interactions in speech-in-noise recognition: Effects of prosodic manipulation as a function of language experience", Journal of the Acoustical Society of America.

[8] Stibbard, R. M., and Lee, J. I. (2006). "Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis," Journal of the Acoustical Society of America 120, 433-442.

[9] Bench, J., Kowal, A. and Bamford, J. (1979). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children", British Journal of Audiology 13, 108-112.