Second-language experience and speech-in-noise recognition: the role of talker-listener accent similarity

Melanie Pinet, University College London, UK

In collaboration with Paul Iverson and Mark Huckvale, University College London, UK

Background

- Speech perception in noise depends on a talker-listener accent interaction.
- Listeners are better at accents that match their own:
  - L1 listeners have a noticeable advantage with L1- over L2-accented speech in speech in noise,
  - L2 listeners can actually be better with L2 speech, particularly when both the talker and the listener share the same L1.
- This accent interaction is well established , but it is still unclear why it occurs.

Background

- **Accent familiarity / L2 experience?**
  - Faster adaptation to familiar L1 accents than unfamiliar L1 or L2 accents (e.g., Adank *et al*, 2009).
  - L2 experience modulates whether L2 listeners have an advantage for L1 or L2 speech (e.g., van Wijngaarden *et al,* 2002, Pinet and Iverson, under revision).
- **Interlanguage benefit?**
  - Intelligibility is enhanced between L2 speakers sharing the same L1.
  - L2 speech mutually more intelligible to L2 listeners (e.g., Bent and Bradlow, 2003).
- **Or acoustic similarity?**
  - The interaction could be driven by the acoustic similarities in the accents of the talkers and the listeners.
  - L2 speakers with acoustically similar accents may be mutually intelligible (e.g., Bent and Bradlow, 2003, vs Stibbard and Lee, 2006).
  - Acoustic similarities in the L1 talker's and the experienced L2 listeners' accent could enhance intelligibility.

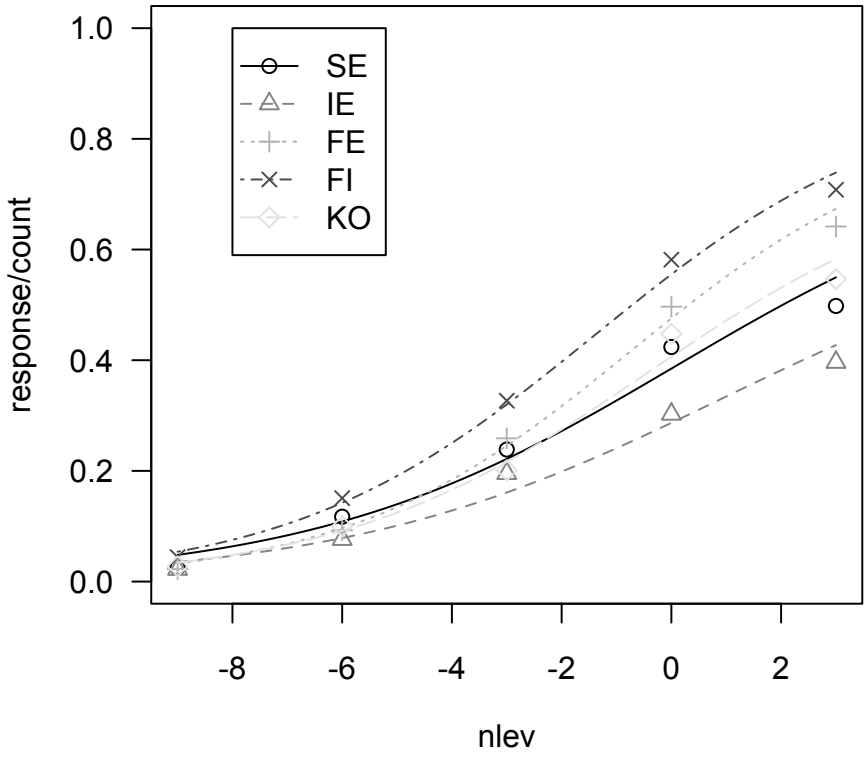
Aims of the study

- Investigate how the acoustic similarity in the talker's and the listener's accent can account for the L1-L2 accent intelligibility in noise
- Explore the impact of L2 experience on this interaction
- Find a reliable measure of accent intelligibility

Method

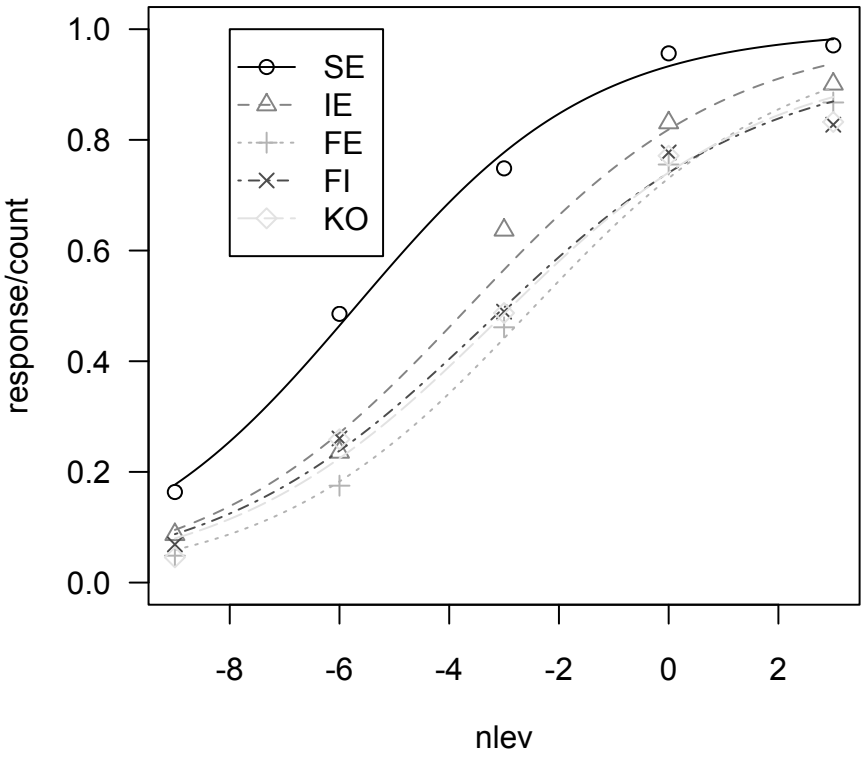
- Listeners: 21 Monolingual Southern British English (SE), 16 French-English Bilinguals (FB), 24 L1 French Experienced (FE), 32 L1 French Inexperienced (FI).
- Talkers: 2 males and 2 females of each accent were recorded reading the BKB sentences:
  - Southern British English (SE)
  - French Experienced (FE)
  - French Inexperienced (FI)
  - Irish English (IE)
  - Korean-accented English (KO)
- The recordings were embedded in speech-shaped noise generated for each individual talker with -9, -6, -3, 0 and +3dB SNR ratios.
- Sentence recognition task on the 5 accents in noise.
- Acoustic analysis using ACCDIST (Huckvale, 2004, 2007a,b).

Results: Speech in noise recognition French Inexperienced Listeners



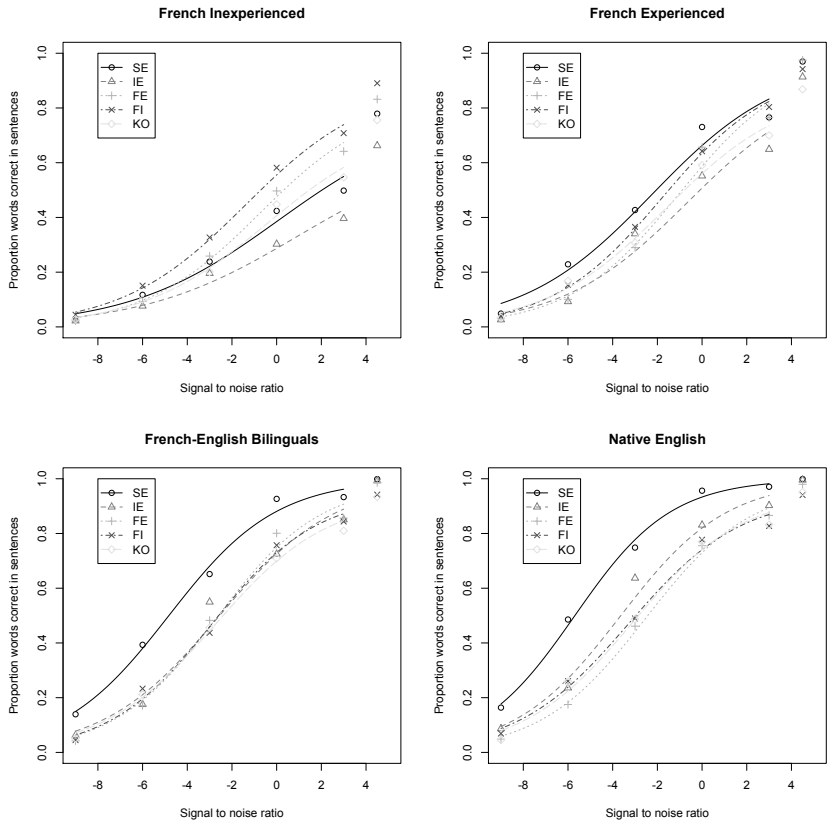
- The FI listeners were most accurate at recognizing FI speech, followed by FE speech.
- Intelligibility became gradually poorer for the other accents
- They displayed graded levels of recognition accuracy.

Results: Speech in noise recognition
Southern British English Listeners



- The SE listeners were most accurate at recognizing SE speech.
- IE speech was only marginally more intelligible than the L2 accents.
- Thus the listeners were selectively tuned to their own accent.

Results: Speech in noise recognition: All groups

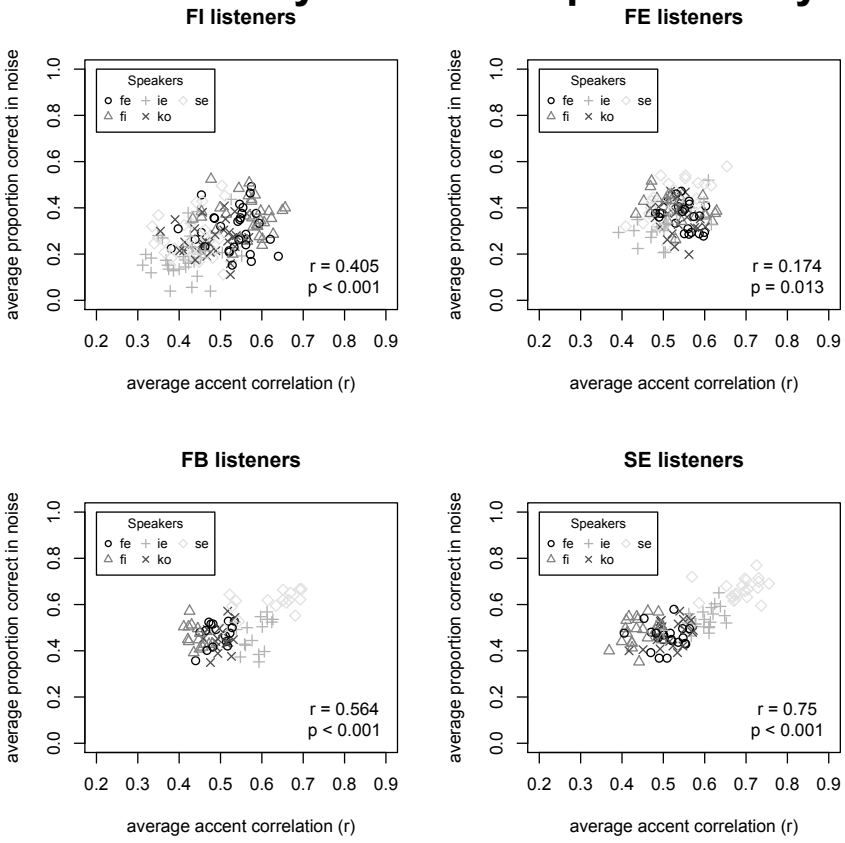


- The more experienced L2 listeners were better at SE speech in noise.
- They thus became selectively tuned to it as their experienced with L1 speech increased.
- The interaction was strongly modulated by the listeners' L2 experience.

Acoustic analysis

- ACCDIST (Huckvale, 2004, 2007a, b): a computational accent quantification method based on making acoustical comparisons of speech produced by pairs of individual talkers.
- Procedure:
  - A SE phonemic transcription is forced aligned against speech recordings of the speakers to segment them.
  - Acoustic measurements are automatically made on the segments (vowel spectra, duration).
  - Segments are compared to each other to create a table of phonetic similarities for each speaker.
  - The assessment of phonetic similarity within each talker removes the influence of global speaker characteristics leaving the phonetic differences that are more indicative of accent.
  - These matrices of within-speaker segmental acoustic distances are then correlated between pairs of talkers.
- Accent distances were correlated with the subjects' identification scores in noise on the different accents.

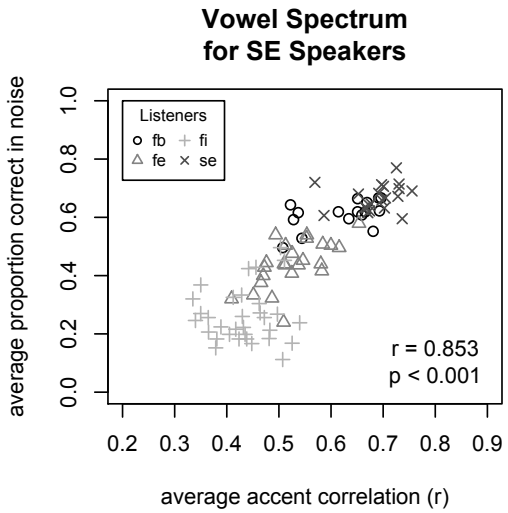
Acoustic analysis: Vowel spectra analysis by listener group

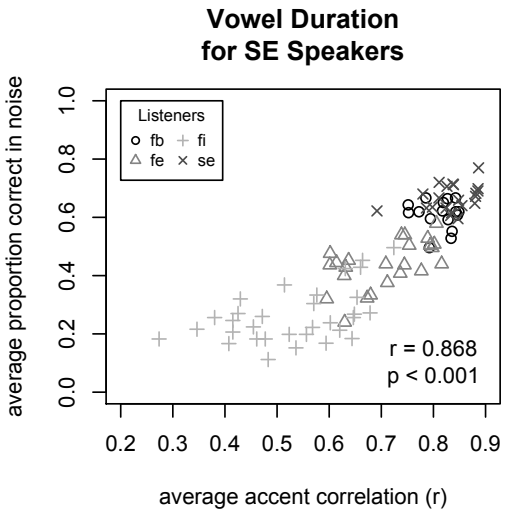


- Speech recognition in noise was highly correlated with accent similarity.
- The listeners were more accurate at recognizing the speech of talkers whose accents closely matched their own acoustically.
- As acoustic distances between accents got wider, the accuracy of speech recognition in noise decreased accordingly

Acoustic analysis by speaker group: Southern British English speakers

- The analysis for the SE talker group shows clear patterns of accent similarity between talkers and listeners as well as displaying the variation in language experience between listener groups.





Conclusions

- L1 background and L2 experience modulate the talker-listener interaction by affecting the listeners' tuning processes in noise (i.e. graded vs selective accent tuning).
- Much of the variance in the relative intelligibility of the different accents could be accounted for in terms of acoustic similarity in the accents of the talker and the listener:
  - Listeners were more accurate at recognizing accents that acoustically matched their own.
  - As acoustic distances between accents got wider, the accuracy of speech recognition in noise decreased accordingly.
  - Acoustic similarity accounts for interactions that accent familiarity or the interlanguage benefit can't justify for.

References

- 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.
- Bent, T., and Bradlow, A. R. (2003). "The interlanguage speech intelligibility benefit," Journal of the Acoustical Society of America 114, 1600-1610.
- Huckvale, M. (2004). “ACCDIST: a metric for comparing speakers’ accents”, Proceedings of the International Conference on Spoken Language Processing, Jeju, Korea, Oct 2004.
- Huckvale, M. (2007a). “Hierarchical clustering of speakers into accents wit the ACCDIST metric”, Proceedings of ICPhS XVI, Saarbrucken, 6-10 August 2007.
- Huckvale, M. (2007b). “ACCDIST: an accent similarity metric for accent recognition and diagnosis” in Speaker Classification II, ed C.Muller, Springer LNAI 4441.
- 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.
- 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.