Listeners' asymmetries in pitch salience judgement of high and low accents

Hae-Sung Jeon[1] and Antje Heinrich[2]
[1]School of Humanities, Language and Global Studies, University of Central Lancashire, UK
[2]Manchester Centre for Audiology and Deafness, University of Manchester, UK

The measurement of intonational targets in spoken utterances is an essential part of intonational analysis. The targets are often realised as local peaks in the fundamental frequency (F0) contours in English. However, the perception of intonational salience is not simply determined by the F0 maximum of the peak but also affected by the precise shape of the F0 contour (Barnes, et al. 2012). For instance, a plateau-shaped F0 peak sounds higher than a sharp peak with the same maximum F0 (Knight, 2008), although it is not clear to what extent this effect is generalised. This study investigated the perception of relative height between two high or low accents in utterance-length speech and non-speech stimuli in different registers.

The stimuli were created to manipulate the following factors: 2 plateau duration in the second accent [25 ms, 100 ms], 2 accent types [high accents (HH), low accents (LL)], 2 registers [high (200-302 Hz), low (132-200 Hz)], and 2 stimulus types [English speech, complex tone]. All factors were crossed. The speech stimuli were resynthesized from a female speaker' speech ('Does Nellie know Lenny?'). Complex tones were harmonic complexes with energy between 200 Hz and 6000 Hz. Each pair of a speech stimulus and a complex tone shared duration and F0 contours. Each listening stimulus had either two high accents (HH) or two low accents (LL). The first accent formed a 25 ms plateau and its height was constant in all stimuli (5.16 ST from the baseline, Figure 1). The second accent height varied in five 1 ST steps between 2 ST below and 2 ST above the first accent. The main outcome measure was the Point of Subjective Equality (PSE), i.e., the height difference between two accents at 0.5 probability indicating the responses at chance level, for each experimental condition. Native British English speakers (N = 57, mean age = 24) using headphones judged which accent sounded higher for HH or lower for LL. Gorilla Experiment Builder was used to host the experiment (Anwyl-Irvine, et al, 2019).

Psychometric functions were fitted to the response data and PSEs were estimated using QuickPsy package with R ver. 3. 6. 1. In addition, mixed effect logistic models were fitted to the response data to estimate the 'second accent' response probabilities using lme4 package. The final model included significant factors based on likelihood ratio tests and listener as a random intercept. The results showed that, first, the longer plateau decreased the PSEs, i.e., increased the perceived salience of the second accent, in all experimental conditions (plateau:100 ms, est. = 0.48, $z = 0.03$, $p < 0.001$). Second, the PSEs tended to be higher for LL than for HH, presumably because listeners were biased to perceive the first accent more salient for LL (accent type: high, est. = 0.32, $z = 0.06$, $p < 0.001$). However, plateaux of 100 ms in the high register speech showed the opposite pattern with lower PSEs for LL, thereby indicating a strong saliency-enhancing effect of the long plateau (stimulus type: tone × accent type: high, est. = -0.25, $z = 0.69$, $p < 0.01$). Third, register (register: low, est. = -0.19, $z = 0.05$, $p < 0.001$) interacted with accent type (register: low × accent type: high, est. = 0.28, $z = 0.07$, $p < 0.0001$), possibly because listeners' sensitivity was reduced for LL particularly in the low register (see Fig. 2).

To summarise, the same magnitude of F0 change in semitones in different directions and registers led to different perceptual consequences. In particular, listeners' reduced sensitivity for low accents poses a challenge in modelling the relationship between the physical F0 and listeners' pitch salience perception. The stimulus type effect showed that listeners' psychoacoustic sensitivity may be modulated by the acoustic complexity of speech and their linguistic expectation.

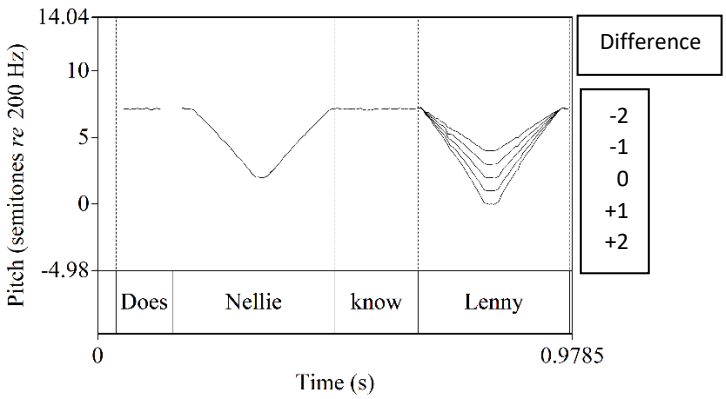

Figure 1: Example F0 tracks of the resynthesised speech stimuli with two low accents (LL) in the high register. Both accents are plateau-shaped (25 ms in this example). The constant F0 at the beginning, between the two accents and at the end is referred to as the baseline. The F0 contour of the HH stimuli is a mirror image of the LL counterpart.

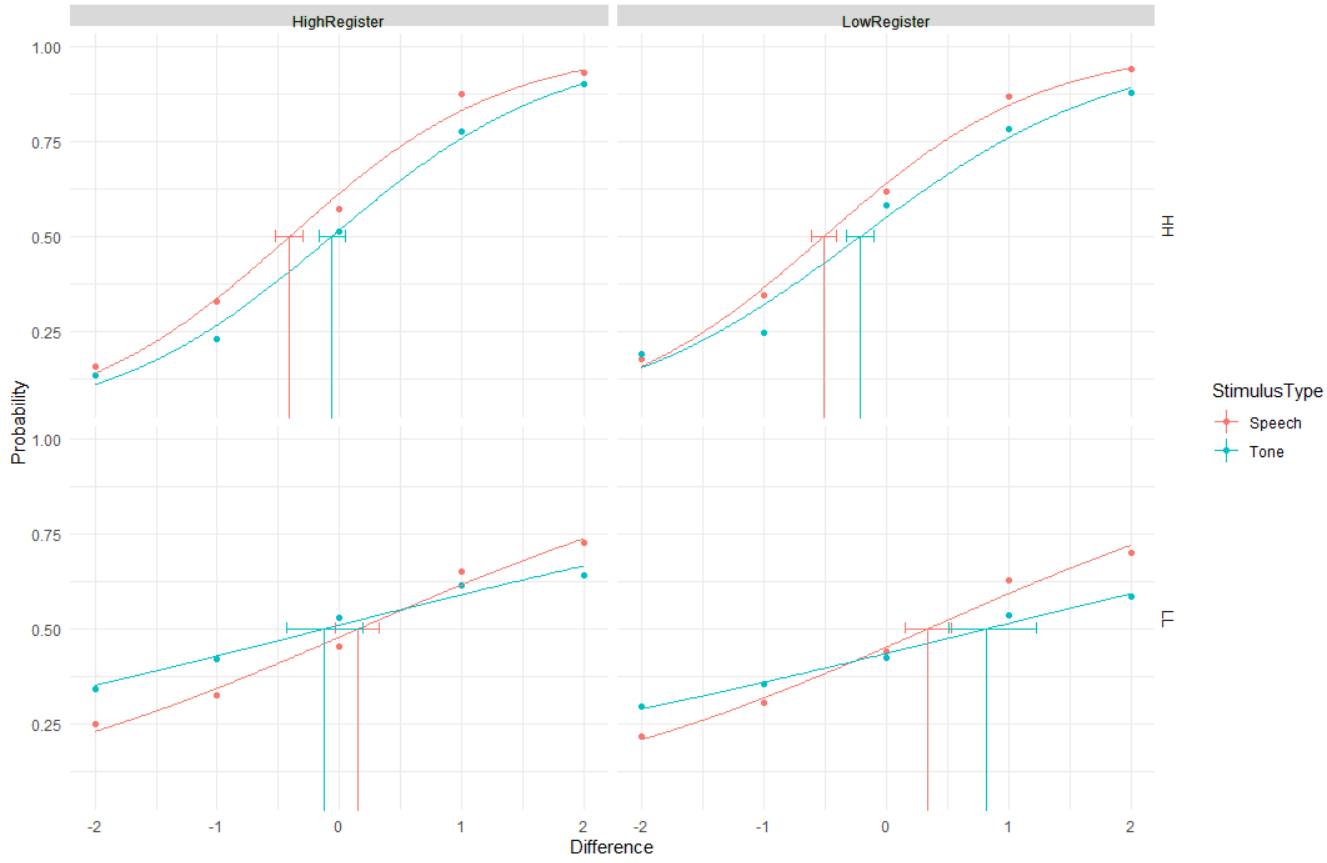

Figure 2: Psychometric functions with averaged probabilities, PSEs and the 95 % confidence intervals. Data were collapsed across second accent plateau duration (25 ms, 100 ms) which did not interact with other factors.

References

Anwyl-Irvine, et al. (2019) *Behavior research methods*, 1-20.

Barnes, J., et al. (2012) In O. Niebuhr (Ed.), *Prosodies: Context, Function, Communication* (pp. 93–118). Berlin/New York: de Gruyter.

Knight (2008) *Lang Sp*, 51, 223–244.