MISMATCH BETWEEN THE PRODUCTION AND PERCEPTION OF F_0 IN NEW ZEALAND ENGLISH ETHNOLECTS

Anita Szakay
Dept. of Linguistics, University of British Columbia, Totem Field Studios, BC, Canada, V6T 1Z4
szakay@interchange.ubc.ca

1. INTRODUCTION

Previous production experiments demonstrated that the two main ethnic dialects of New Zealand English significantly differ in prosodic features, such as rhythm and intonation (e.g. Britain 1992, Warren 1998, Szakay 2006). Szakay (2006) also showed that Maori English mean pitch is significantly higher than that of Pakeha English, the dialect spoken mainly by speakers of European descent. The study also indicated that the increasing F_0 values of Maori English are a result of a change in progress, with young Maori speakers using significantly higher mean pitch than young Pakeha speakers. Fig. 1. – taken from Szakay (2006) – demonstrates the interaction between age and ethnicity as predictors of mean pitch values ($p<.05$) in New Zealand.

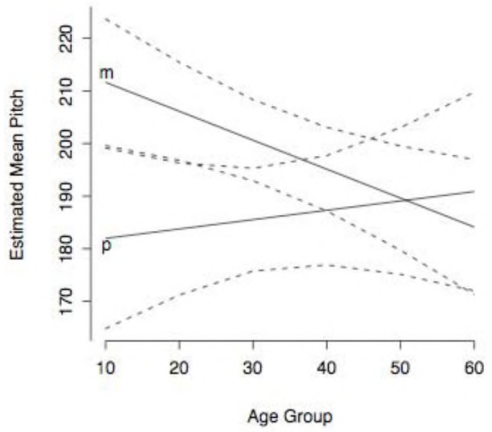

Fig. 1. Regression model predictions for estimated mean F_0 (Hz) for Maori speakers (m) and Pakeha speakers (p). Values adjusted to female speakers. Dashed lines indicate 95% confidence intervals.

The present perception study is part of a larger research project using innovative techniques to isolate the precise prosodic features that listeners might tune into in ethnic dialect identification in the New Zealand context. The study set out to investigate whether naïve listeners are aware of the differing suprasegmental features of the two ethnic dialects and whether they can identify the dialects if there are no segmental cues available in the speech signal. The results reported here relate to the perception of mean pitch only. For an extensive coverage of the perceptual relevance of other prosodic cues consult Szakay (2008).

2. METHOD

107 listeners (52 Maori, 55 Pakeha) performed a forced-choice dialect identification task. Based on the speech samples of 20 speakers (10 Maori, 10 Pakeha), seven speech conditions were created, each keeping different suprasegmental information in the speech signal (e.g. low-pass filtered condition, resynthesized rhythm-only at mean pitch condition, intonation-only condition). Condition Three was created as a monotonous speech rhythm only condition. Each consonant and pause was replaced by silence, while vowels were replaced by a tone complex created in Praat as a sum of a number of cosine waves with equidistant frequencies at a sampling frequency of 8000Hz. It was created at the mean pitch across all speakers according to gender (118Hz for males, 188Hz for females). A sample spectrogram of Condition Three is given in Fig. 2.

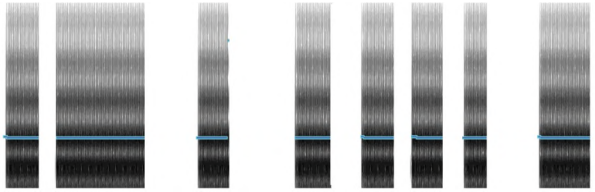

Fig. 2. Sample spectrogram for Condition Three: Rhythm only at mean pitch across speakers.

3. RESULTS

The results of logistic regression analyses in each condition indicate that listeners do rely on the F_0 characteristics to identify speaker ethnicity. In particular, Condition Three revealed a significant interaction between participant ethnicity and speaker mean pitch as predictors of perceived speaker ethnicity ($p<.05$). Fig. 3. illustrates the predictions of the linear regression model, where the y-axis shows the probability of perceived speaker ethnicity as Pakeha. Higher values indicate a Pakeha response, while low values indicate a Maori response. Pakeha participants identify a speaker with high mean pitch as Pakeha, and a speaker with low pitch as Maori. Maori participants do not rely on mean pitch in this condition. This result is at odds

with the production results, where Maori speakers in fact produce significantly higher mean F_0 values.

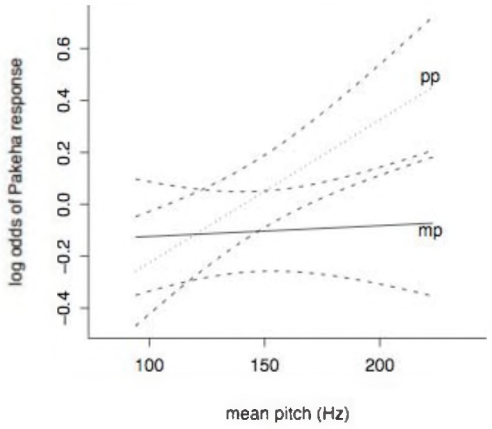

Fig. 3. Regression model predictions of perceived speaker ethnicity by speaker mean pitch and participant ethnicity (pp=Pakeha participant, mp=Maori participant). Dashed lines indicate 95% confidence intervals.

4. DISCUSSION

The results indicate that participants do rely on the mean pitch of a speaker in Condition Three, where many of the other cues have been eliminated from the speech signal. However, they perceive a lower mean pitch to be a characteristic of a Maori speaker and a higher mean pitch to indicate a Pakeha speaker. This is completely different from what is happening in production, where Maori speakers in fact use a significantly higher mean pitch. As this seems to be a change over time, it is plausible that listeners are not yet consciously aware of this new feature of Maori English. When people are overtly asked whether they think Maori speakers have a higher or lower mean pitch than Pakeha speakers, they tend to reply 'lower pitch' without hesitation. This might be the result of certain physical stereotypes held about Maori being big and bulky. However, when they are asked to imitate a Maori speaker, they almost always use a higher pitch in doing so. This suggests that subconsciously they might be aware of the ongoing change in Maori English pitch and, if such is the case, then we might expect that with time, perception results regarding mean pitch will adjust as listeners become actively aware of this new prosodic feature of the ethnolect.

This suggests that stereotypes might have a stronger effect on speech perception than the actual change in progress in this ethnolect of New Zealand English. Perception studies in the US yielded similar results with regards to the pitch characteristics of a speaker, where lower F_0 levels were associated with African Americans and higher F_0 with European Americans (Hawkins 1992, Foreman 2000, Thomas & Lass 2005).

Accounting for stereotypes – or ideologies – about certain speaker groups causes problems for current exemplar theoretic frameworks of speech production and perception. In an exemplar based model, categories are made up of a large set of remembered exemplars from a wide range of speakers, and the auditory properties that distinguish speakers are retained in these exemplars (e.g. Pierrehumbert 2001, Johnson 1997). The fact that stereotypes can get activated in speech perception instead of actually encountered exemplars with stored phonetic and social detail suggests that the framework should be revised accordingly.

REFERENCES

Britain, D. (1992), 'Linguistic Change in Intonation: The Use of High Rising Terminals in New Zealand English', *Language Variation and Change* **4**(1), 77–104.

Foreman, C. (2000), 'Identification of African-American English from Prosodic Cues', *Texas Linguistic Forum* **43**, 57–66.

Hawkins, F. D. (1992), Speaker Ethnic Identification: The Roles of Speech Sample, Fundamental Frequency, Speaker and Listener Variations, PhD thesis, University of Maryland at College Park.

Johnson, K. (1997), Speech perception without speaker normalization: An exemplar model, in K. Johnson and J. W. Mullennix, eds, 'Talker Variability in Speech Processing', San Diego: Academic Press, pp. 145–165.

Pierrehumbert, J. (2001), Exemplar dynamics: Word frequency, lenition and contrast, in J. Bybee and P. Hopper, eds, 'Frequency effects and the emergence of lexical structure', John Benjamins, Amsterdam, pp. 137–157.

Szakay, A. (2006), Rhythm and pitch as markers of ethnicity in New Zealand English, in P. Warren & C. Watson, eds, 'Proceedings of the 11th Australasian International Conference on Speech Science & Technology', University of Auckland, pp. 421–426.

Szakay, A. (2008), *Ethnic Dialect Identification in New Zealand: The Role of Prosodic Cues*. VDM Verlag Dr Müller.

Thomas, E. R. & Lass, N. J. (2005), 'Cues for Identification of African American and European American Speakers', Poster presented at the Annual Convention of the American Speech-Language-Hearing Association.

Warren, P. (1998), 'Timing Patterns in New Zealand English Rhythm', *Te Reo: Journal of the Linguistic Society of New Zealand* **41**, 80–93.

ACKNOWLEDGEMENTS

Special thanks to Jen Hay, Alex D'Arcy, Paul Warren and Margaret Maclagan. The funding provided by the University of Canterbury, New Zealand is also gratefully acknowledged.