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.

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 at 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.

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.

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**


**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.