Volume 14
Issue 2 *Selected Papers from NWAV 36*

Article 18

11-22-2008

Social Networks and the Perceptual Relevance of Rhythm: A New Zealand Case Study

Anita Szakay
University of British Columbia

Social Networks and the Perceptual Relevance of Rhythm: A New Zealand Case Study

Abstract

Previous research has demonstrated that the two main ethnolects of New Zealand English display distinct rhythmic qualities (e.g., Warren 1998; Szakay 2006). Using the normalized vocalic Pairwise Variability Index (PVI) to measure rhythm, as suggested by Grabe and Low (2002), Szakay (2006) showed that Maori English is significantly more syllable-timed than Pakeha English, the main variety used by speakers of European descent. The present study set out to investigate whether listeners are aware of this prosodic difference and whether they are capable of tuning in to speaker rhythm to facilitate dialect identification.

A perception experiment was carried out using 20 speakers and 107 listeners, with nearly equal numbers of Maori and Pakeha participants. As the linguistic experience of participants has been shown to affect the perception of dialect variation (e.g., Preston 1986; Tamasi 2003; Clopper 2004), each listener was assigned a Maori Integration Index (MII) to measure their previous exposure to Maori English.

To isolate the precise features that listeners might attend to when identifying speaker ethnicity, various speech conditions were created, each retaining different prosodic information in the speech signal. Thus, listeners were asked to perform a forced choice ethnolect identification task in the following conditions:

(a) unaltered speech
(b) low-pass filtered speech at 400Hz
(c) resynthesized rhythm and intonation together
(d) resynthesized rhythm only

The results indicate that, overall, New Zealanders are in fact aware of the rhythmic difference between Maori English and Pakeha English, and are capable of attending to speaker rhythm in a dialect identification task, perceiving syllable-timed speakers as Maori-sounding, and stress-timed speakers as Pakeha-sounding. However, not all participants are equally good at exploiting speaker rhythm to facilitate dialect identification. Logistic regression analyses reveal that PVI interacting with MII is a significant predictor of perceived ethnicity in all conditions. The results demonstrate that listeners with a higher MII are significantly better at relying on rhythm than those participants who are less integrated into Maori social networks. These low MII participants only tend to use rhythm correctly in the unaltered and low-pass filtered speech conditions where many other cues are also available. In the more degraded listening conditions they either do not rely on rhythm as a cue or use it incorrectly.

The fact that highly integrated listeners are able to rely on rhythm more accurately than low-MII listeners confirms the hypothesis that greater exposure to a dialect facilitates the identification of not only segmental but also prosodic features as belonging to the particular dialect. The results in general also demonstrate the role of social network not just in terms of frequency of forms but also in terms of accuracy in identifying in-group vs. out-group members.

Cover Page Footnote

Special thanks to Jen Hay, Alex D'Arcy, Margaret Maclagan and Paul Warren for their support and invaluable comments.

Social Networks and the Perceptual Relevance of Rhythm: A New Zealand Case Study

Anita Szakay*

1 Introduction

As the number of fluent, native Māori language speakers has steadily decreased in New Zealand, English has become the dominant language of almost all Māori people, although many are also familiar with the Māori language. In such circumstances, it is not surprising that a distinctive variety of Māori English has emerged to express ethnic identity and positive attitudes toward Māori culture (Holmes, 2005).

The term New Zealand English (NZE) is not as well defined as it first might seem. King (1993) points out that what most sociolinguists would call NZE, is probably better labelled as Pākehā English, that is, the English spoken mainly by European New Zealanders. King also suggests that Māori English is not restricted to ethnically Māori speakers, but is also used by some Pākehā who either grew up or identify with Māori peer groups. It is also the case that not all ethnically Māori speak Māori English.

Previous research has suggested that the differences between Māori English and Pākehā English tend to be relative rather than absolute. There are many features that are shared by both dialects but where the frequency of forms in each variety differs. Since the 1990s, linguists working on Māori English have made numerous attempts to identify the core features that differentiate the two dialects, at least quantitatively if not qualitatively. Some of these studies concentrated on phonological features, such as the pronunciation of vowels and consonants (e.g. King, 1993; Robertson, 1994; Holmes, 1996; Bell, 2000), while others set out to identify possible prosodic differences (e.g. Bauer, 1994; Bell, 2000; Holmes and Ainsworth, 1996, 1997; Warren, 1998).

Although English in general is considered to be stress-timed, it has been suggested that NZE shows a tendency towards syllable-timing. Previous work on rhythm in New Zealand has also claimed that there may be a difference in timing patterns within the variety, with Māori speakers producing more syllable-timed speech than Pākehā speakers (Ainsworth, 1993; Holmes and Ainsworth, 1996, 1997; Warren, 1998).

Using the Pairwise Variability Index (Grabe and Low, 2002), Szakay (2006) investigated the differing rhythmic properties of the two ethnolects and found that Māori English is indeed significantly more syllable-timed than Pākehā English.

The vocalic Pairwise Variability Index (PVI) is based on the relative difference in duration of successive vocalic segments and is normalized for local rate variations. A low PVI value shows less variation in vowel duration, and as such indicates a more syllable-timed language. Stress-timed languages, on the other hand, typically demonstrate shorter unstressed vowels alternating with longer vowels, resulting in a higher PVI. Using PVI as a measurement for syllabic rhythm, it is possible to plot languages along a continuum of stress- and syllable-timing instead of categorically distinguishing between stress- and syllable-timed languages.

The production experiment described in Szakay (2006) also showed that younger speakers in general produce more syllable-timed speech and that social network strength scores correlate with speech rhythm. A Māori Integration Index (see Section 2.7 below) was assigned to each speaker and it was found that highly integrated speakers exhibit more syllable-timed rhythmic patterns than non-integrated speakers.

The production study served as a precursor to the dialect identification experiment, which is the focus of the present paper. The main objective of the perception task was to investigate whether listeners can correctly identify a speaker's ethnicity and to find out what cues listeners use for ethnic identification. Can they identify Māori English and Pākehā English based solely on prosodic features, or do they also need segmental cues? The question also arose whether listeners are aware

*Special thanks to Jen Hay, Alex D'Arcy, Margaret Maclagan and Paul Warren for their support and invaluable comments.

of the dialectal differences shown by the results of the production experiment in Szakay (2006). If they are conscious of the contrasting rhythmic properties between the two dialects, could they accurately use them as cues in the perception task? Another aim of the present study was to investigate whether social network strength scores would also play a role in dialect perception as they did in the production study.

2 Methodology

2.1 Speakers

The previous production experiment analyzed the speech of 36 New Zealanders, for each of whom a reading passage and a narrative were recorded. The present perception study selected 20 of those speakers and concentrated on the narratives only. Our aim was to choose 20 speakers who were easily identifiable as Māori or Pākehā in a normal, unmodified speech condition. We argue that there would be not much point in trying to identify a speaker's ethnicity from suprasegmental cues alone, if they were unlikely to be correctly identified based on unaltered speech. In order to decide which speakers to include in the experiment, a test session was run among linguists in which they rated each speaker based on how Māori or Pākehā-sounding they found them to be. The distribution of the selected speakers according to ethnicity, gender and age is shown in Table 1. The age range for the *younger* groups was 18–30, while for the *older* groups it was 40–60+.

Table 1: Distribution of the 20 speakers used in the perception task according to ethnicity, gender and age.

	younger male	older male	younger female	older female
Māori	4	2	2	2
Pākehā	4	2	2	2

The reason why there were twice as many younger male speakers chosen to be included in the experiment is twofold. On the one hand, it has been previously argued that the features of Māori English are most salient for younger male speakers (King, 1999). Thus, we were more confident that this group would yield positive results in a dialect identification task. On the other hand, we simply had more younger male speakers recorded than any of the other groups.

2.2 Creating the Conditions

In order to investigate exactly what cues listeners tune in to when identifying an accent, seven test conditions were created. Our main aim was to be able to disentangle the various effects of different linguistic cues in order to establish what exactly it means to sound Māori or Pākehā. Each condition contained different linguistic cues available for the listener. Based on these we would try to determine whether rhythm or intonation, alone or combined, could serve as a sufficient cue for identifying the ethnicity of the speaker.

The first six conditions provided only suprasegmental information, either separately or in various combinations. In Conditions One to Five we used speech re-synthesis, while in Condition Six we applied low-pass filtering to eliminate segmental information. The final test condition presented the unmodified version of the passages. The present study concentrates on four of the seven conditions that are relevant to the analysis of rhythm.[1] These conditions are:

(a) Condition Two: Rhythm Only at Mean Pitch
(b) Condition Five: Rhythm and Intonation Together
(c) Condition Six: Low-pass Filtered at 400Hz
(d) Condition Seven: Unmodified

[1]For information on the rest of the conditions consult Szakay (2007).

2.2.1 Condition Two - Rhythm Only at Mean Pitch

The purpose of this condition was to investigate whether participants could identify the two dialects if the one and only available cue they had was rhythm. To create a rhythm only condition, all segmental information was eliminated and the intonation was flattened. All passages had been manually segmented into vowel and consonant sequences and marked up in a Praat textgrid. A Praat script was written that replaced each consonant and each pause by silence, while vowels were replaced by a tone complex, also created in Praat. This technique ensured that what listeners would hear as rhythm in the perception task, would closely correspond to the vocalic PVI values assigned previously to each speaker in a production experiment (Szakay, 2006). The sound replacing vowels was produced from a tone complex 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. This was 118Hz for male speakers and 188Hz for female speakers.[2]

Note that previous research used a different method for creating a rhythm only condition. For example, Leyden (2004) first low-pass filtered the samples to obtain a rhythm and intonation condition, then the speech was monotonized by changing the pitch contour into a flat line. However, after such transformation the acoustic signal becomes extremely degraded. Gooskens and van Bezooijen (2002) omitted this condition in their study for similar reasons.

Figure 1 demonstrates a sample spectrogram[3] from the flat rhythm only condition. When compared with the corresponding unmodified version shown in Figure 4, it can be observed how the segments marked as consonants were replaced by silence. The straight line of the pitch tracker in Figure 1 illustrates that the intonation is flattened. In this sample, the pitch is kept constant at 118Hz, as for all male speakers.

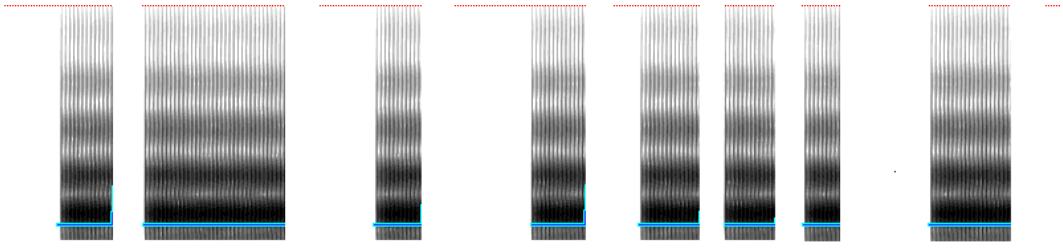

Figure 1: Sample spectrogram from Condition Two (rhythm only)

2.2.2 Condition Five - Rhythm and Intonation

In Condition Five our aim was to find out how accurately participants would identify the two dialects if they could hear both rhythm and intonation together. First a hummed version of each sound file was created. Each original vocalic segment was then replaced by the corresponding hummed segment, while consonants and pauses were replaced by silence. This condition is similar to Condition Two, with the exception that all vowels retain their original pitch movements instead of being replaced by a monotonous tone complex.

Compare the spectrogram for this condition, shown in Figure 2, with the spectrogram taken from Condition Two (Figure 1). The only difference between the two is that the pitch dynamics during vocalic segments are now clearly visible, as opposed to Condition Two, where intonation was flattened.

[2]Gender information was kept in the signal so it could be investigated whether rhythm is interpreted differently in the case of male and female voices.

[3]All spectrograms shown in this section represent the same speech fragment from one male speaker.

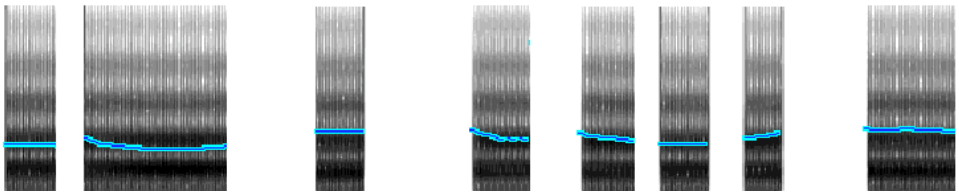

Figure 2: Sample spectrogram from Condition Five (rhythm and intonation)

2.3 Condition Six - Low-pass Filtered at 400Hz

Several previous studies have used low-pass filtered speech for dialect identification tasks (e.g. Foreman, 2000; Frota et al., 2002; Leyden, 2004; Thomas and Reaser, 2002). It is claimed that low-pass filtering at 400Hz eliminates segmental information, which is mainly contained in the higher frequencies. At the same time low-pass filtering is said to retain both syllabic rhythm and intonation, as pitch rarely rises higher than 400Hz. However, listening to filtered speech makes it obvious that it is not only rhythm and intonation that are preserved. In addition to rhythm and intonation, low-pass filtering seems to also retain the voice quality of the speech. This makes Condition Six somewhat different from Condition Five, and accordingly both were used in the perception task. If participants' accuracy differs in the two conditions, it would suggest that they are using voice quality as a cue in dialect identification.[4][5]

Low-pass filtering at 400Hz was carried out using Praat with a smoothing of 50Hz. As the filtering produces a muffled sound, the amplitude was multiplied by four to increase loudness. This ensured that Condition Six would not be noticeably quieter than the other conditions. Figure 3 shows how all spectral information above 400Hz is eliminated from the speech signal.

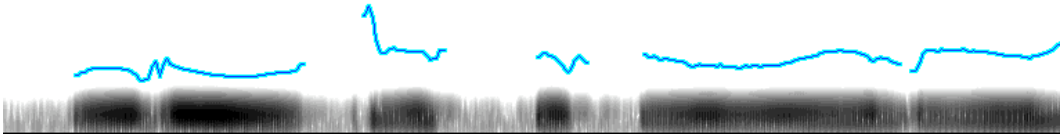

Figure 3: Sample spectrogram from Condition Six (LP-filtered at 400Hz)

2.3.1 Condition Seven - Unaltered speech

The final condition contained the original passages and acted as a control condition. Although this is the unaltered condition, two changes were nonetheless carried out. To make sure that participants would not base their judgements on non-standard syntax rather then on the actual phonetic properties of the speech, two words were cut from two of the original passages. The word *more* was deleted from the phrase *more faster* (speaker *m02*), while *don't* was deleted from the phrase *you don't see nothing* (speaker *m28*). A test run was carried out among a group of linguists to make sure that the altered phrases still sounded natural.[6] A sample spectrogram with its corresponding textgrid from the unmodified speech condition is shown in Figure 4.

[4] And possibly some limited segmental information, such as F1, for example.

[5] A follow-up study is being carried out that aims to investigate possible voice quality (=phonation) differences between the two ethnolects.

[6] These words were present for the other conditions, as Condition Seven was dealt with after the speech resynthesis had been completed. They were also taken into account when PVI values were calculated in the production experiment.

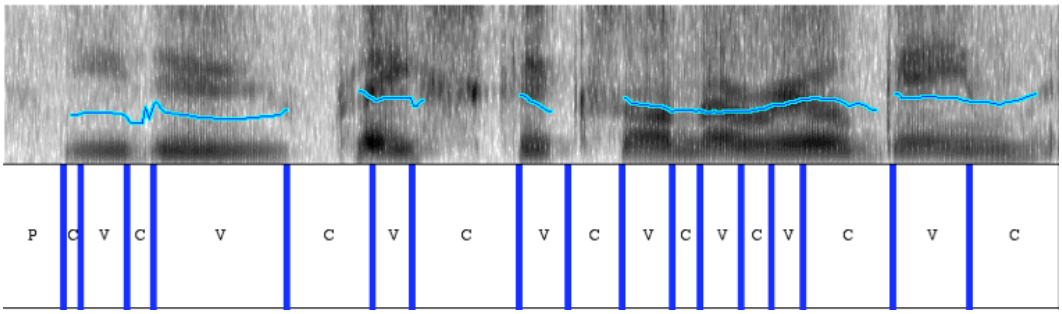

Figure 4: Sample spectrogram from Condition Seven (unmodified speech)

2.4 Stimulus Tape

The seven speech conditions were organized into seven blocks with each of the twenty passages, one from each speaker, randomized within each block. The various test conditions were played in an ascending order, starting with Condition One. That is, the presentation of material was so arranged that the amount of linguistic information that was made available to the listeners increased from one block to the next one. This prevented listeners from transferring information gathered from one block over to the next block.

The mean duration of the passages used in the task was 13.4 seconds, with the actual lengths ranging between 10–15 seconds. As the task was quite lengthy, a two minute break was held after Condition Four. Thus, Part One consisted of 80 passages in four conditions and Part Two consisted of 60 passages in the remaining three conditions. All in all, each experimental session lasted about 55 minutes, including instruction time and the filling out of a background information sheet.

2.5 Answer Sheets

Participants were issued with response sheets on which they were asked to indicate, for each passage, whether they thought a particular speaker was Māori or Pākehā. They responded by circling a number on a four-point scale ranging from 1 *'very Pākehā sounding'* to 4 *'very Māori sounding'*. Participants were provided with information relating to each speaker's age (*'young'* or *'old'*) and gender; they had only to decide on the speaker's perceived ethnicity.

Participants were required to always circle a scale position, even if they felt they had to guess. It should be noted that, since the scale has no midpoint, the subjects were forced to make a decision, no matter how tentative. Although a four point scale was presented on the answer sheet, the statistical analysis treated the perceived ethnicity of the speaker as a strictly binary variable: a score of 1 or 2 = Pākehā, a score of 3 or 4 = Māori.

2.6 Listeners

Altogether, 107 participants took part in the perception experiment, all born and raised in New Zealand. 55 of them marked themselves as Pākehā and 52 claimed to be of Māori descent. The distribution of the participants according to age, gender and ethnicity is shown in Table 2 .

2.7 Social Network Strength Scores: Māori Integration Index

Previous research (e.g. Clopper, 2004) has indicated that people with more exposure to a particular dialect are more accurate in identifying that dialect in a perception experiment. In order to test this hypothesis in the New Zealand context, it was necessary to devise a procedure for characterising social network structure which reflects local social practice. A Māori Integration Index (MII) was designed to measure participants' level of involvement in Māoridom. The MII was constructed form responses to eight questions presented on the background information sheet. Points could be scored

Table 2: Distribution of the 107 participants in the perception task according to ethnicity, gender and age.

	10's	20's	30's	40's	50's	60's
Māori female	5	8	5	2	2	2
Māori male	5	11	5	3	2	2
Pākehā female	9	12	2	4	3	3
Pākehā male	6	3	4	5	2	2

for own ethnicity, spouse's ethnicity, competence in the Māori language and general involvement in things Māori. Questions also related to how often the speaker watches Māori Television or listens to local iwi (tribe) radio stations or visits a marae (a quintessential gathering place for the Māori community) Subjects were also asked about the ethnicity of the people they spend most of their time with, either at work or in their spare time. They also had a chance to state to what extent they perceive themselves to have been exposed to Māori English.

Two slightly different scales for Māori and Pākehā subjects were used, with a Pākehā participant being able to score slightly more points for the same answer than a Māori participant. In New Zealand culture, which is a predominantly Pākehā English environment, for a Pākehā participant it possibly takes more of a conscious effort to be involved with the Māori community than for a Māori subject, who is often intrinsically involved. Based on this, Pākehā participants could potentially score an extra half point for each question compared to Māori participants. This excludes the question on the participant's own ethnicity where Pākehā subjects score zero. This way the maximum possible score was 16 for Māori and 17.5 for Pākehā subjects.

The histograms in Figure 5 illustrate the distribution of the Māori Integration Indices amongst Māori participants (left panel) and Pākehā participants (right panel). The average score for Māori participants was 11.95, with the median at 13, while Pākehā averaged a score of 3.3 with their median being at 2.5.

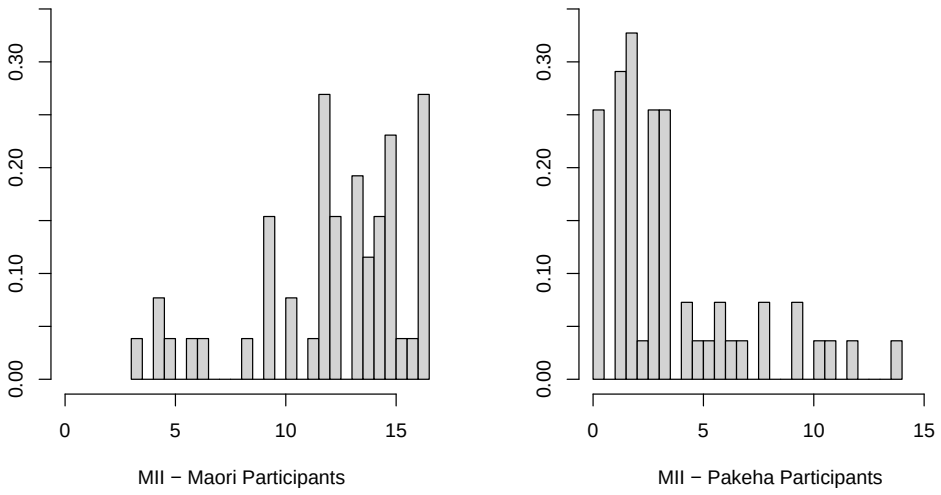

Figure 5: Histograms indicating the distribution of Māori Integration Indices for Māori participants (left panel) and Pākehā participants (right panel).

In Szakay (2006) none of the Pākehā speakers scored higher on the MII than any Māori speaker,

making it difficult to tell whether the MII was doing more than just separating the two ethnic groups. In the perception experiment, however, the distribution of MII scores amongst participants was very different, with the desired result of some Pākehā scoring higher than some Māori subjects. That is, some Pākehā listeners were more integrated into Māori society than some low-scoring Māori participants. This provides a good testing ground to examine whether the MII is a better indicator of participants' accuracy in the dialect identification task than the social variable of *ethnicity* alone.

3 Results and Discussion

A logistic regression model was fit by hand for each condition which tested the effects on perceived speaker ethnicity. Overall, the results of the perception experiment clearly indicate that participants are able to attend to the rhythmic characteristics of a speaker and use them as a cue for distinguishing between the two ethnic dialects of New Zealand English. Previous research has suggested that there is a difference in timing patterns between Māori English and Pākehā English, however, no other study has shown that naive listeners are in fact aware of the variation and are capable of tuning into this rhythmic difference to help identify a speaker's ethnicity. The statistical analysis reveals that PVI is a significant predictor of perceived ethnicity in all relevant conditions ($p<.001$), suggesting that listeners can and do make use of a speaker's rhythmic properties even in degraded listening conditions to facilitate dialect identification.

However, not all participants are equally good at using PVI the correct way. The interaction of speaker rhythm and participants' MII in Conditions Seven, Six, Five and Two shows that highly integrated listeners are predictably better at interpreting PVI values to correctly identify a speaker's ethnicity ($p<.01$ in all four conditions). These highly integrated listeners always perceive a stressed-timed speaker as more Pākehā sounding, and a more syllable-timed speaker as more Māori sounding. This is in line with the results of the production experiment described in Szakay (2006), which demonstrated that Māori English speakers are in fact significantly more syllable-timed than Pākehā speakers. Listeners with a low MII, on the other hand, only manage to use PVI in the right direction in the unaltered speech and the low-pass filtered speech conditions. In the rhythm and intonation only condition they are not able to rely on rhythm, while in Condition Two they use it in an opposite direction. This difference in the perception of syllabic rhythm between highly and lowly integrated listeners indicates that greater exposure to a dialect not only facilitates listeners' ability to recognise and identify segmental characteristics of particular vowels and consonants but also the prosodic features of the particular dialect.

Figure 6 summarizes the interaction between speaker rhythm and participant MII in the four relevant conditions. The y-axis of these 3D diagrams shows the likelihood of a Pākehā response by listeners. A higher value indicates that the speaker is more likely to be identified as Pākehā, while a low value signals a Māori response. The x-axis demonstrates the continuum from syllable-timing to stress-timing with the increasing PVI values. The predicted responses of highly integrated participants are shown in the back panel, while the front panel of the diagram displays the perceived ethnicity by listeners with a low MII.

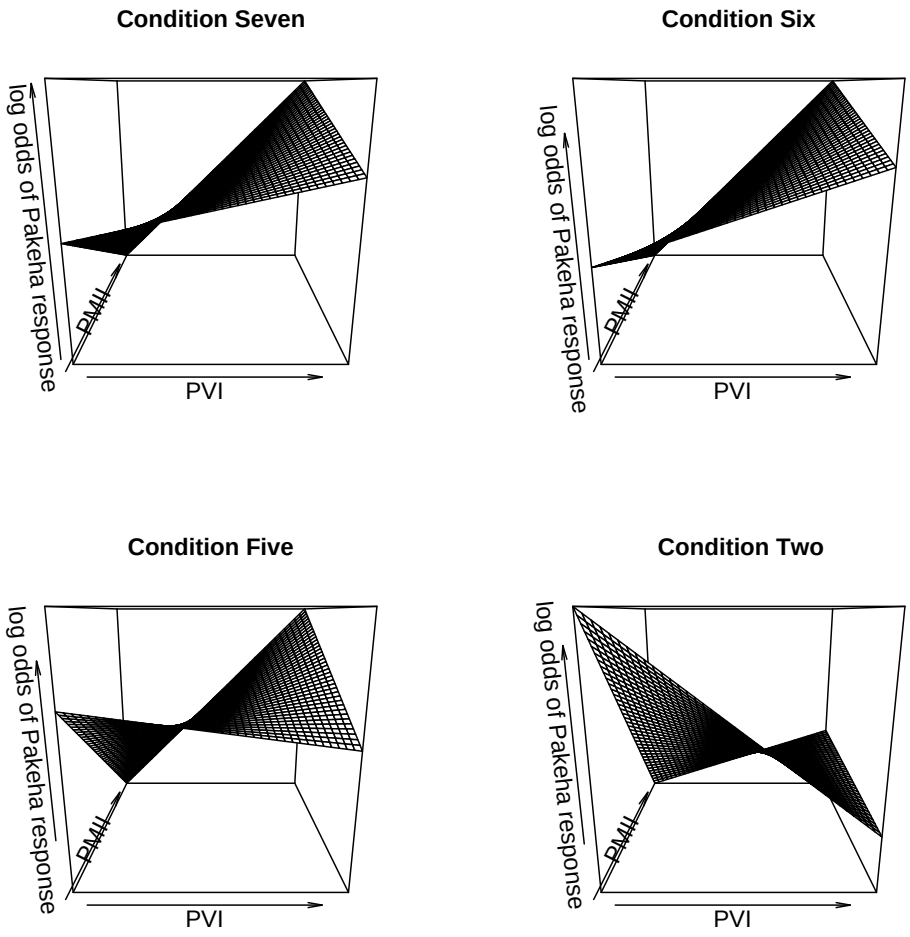

Figure 6: Summary of the interaction between participant MII and speaker rhythm in the models of perceived ethnicity in Conditions Seven, Six, Five and Two.

4 Conclusion

The present study reveals that naive New Zealand listeners are aware of the rhythmic difference between Pākehā English and Māori English and are capable of making use of speech rhythm when identifying a speaker's ethnicity. The fact that highly integrated listeners are able to rely on rhythm more accurately than low-MII listeners confirms the hypothesis that greater exposure to a dialect facilitates the identification of not only segmental but also prosodic features as belonging to the particular dialect. The results in general also demonstrate the role of social network not just in terms of frequency of forms but also in terms of accuracy in identifying in-group vs. out-group members.

References

Ainsworth, Helen. 1993. Rhythm in New Zealand English. Ms., Victoria University of Wellington.

Bauer, Laurie. 1994. English in New Zealand. In *The Cambridge History of the English Language*, ed. Robert Burchfield, volume 5, 382–429. Cambridge: Cambridge University Press.

Bell, Allan. 2000. Maori and Pakeha English: A case study. In *New Zealand English*, ed. A. Bell and K. Kuiper, 221–248. Wellington: Victoria University Press.

Clopper, Cynthia G. 2004. Linguistic Experience and the Perceptual Classification of Dialect Variation. Doctoral Dissertation, Indiana University.

Foreman, Christina. 2000. Identification of African-American english from Prosodic Cues. *Texas Linguistic Forum* 43:57–66.

Frota, Sonia, Marina Vigário, and Fernando Martins. 2002. Language Discrimination and Rhythm Classes: Evidence from Portuguese. In *Proceedings of the 1st International Conference on Speech Prosody*, 315–318. Aix-en-Provence, France.

Gooskens, Charlotte, and Renée van Bezooijen. 2002. The role of prosodic and verbal aspects of speech in the perceived divergence of Dutch and English language varieties. In *Present-day Dialectology. Problems and Findings*, ed. Jan Berns and Jaap van Marle, 173–192. Berlin: Mouton de Gruyter.

Grabe, Esther, and E. Low. 2002. Durational Variability in Speech and the Rhythm Class Hypothesis. In *Papers in Laboratory Phology*, ed. C. Gussenhoven and N. Warner, volume 7, 515–546. Mouton de Gruyter.

Holmes, Janet. 1996. Losing voice: is final /z/ devoicing a feature of Maori English? *World Englishes* 15:193–205.

Holmes, Janet. 2005. Using Māori English in New Zealand. *International Journal of the Sociology of Language* 172:91–115.

Holmes, Janet, and Helen Ainsworth. 1996. Syllable-timing and Maori English. *Te Reo: Journal of the Linguistic Society of New Zealand* 39:75–84.

Holmes, Janet, and Helen Ainsworth. 1997. Unpacking the research process: investigating syllable-timing in New Zealand English. *Language Awareness* 6:32–47.

King, Jeanette. 1993. Maori English: a phonological study. *New Zealand English Newsletter* 7:33–47.

King, Jeanette. 1999. Talking Bro: Māori English in the University Setting. *Te Reo: Journal of the Linguistic Society of New Zealand* 204:20–38.

Leyden, K. van. 2004. Prosodic Characteristics of Orkney and Shetland Dialects. An Experimental Approach. Doctoral Dissertation, Leiden University.

Robertson, Shelley. 1994. Identifying Maori English: A study of ethnic identification, attitudes, and phonetic features. Master's thesis, Victoria University of Wellington, New Zealand.

Szakay, Anita. 2006. Rhythm and pitch as markers of ethnicity in New Zealand English. In *Proceedings of the 11th Australasian International Conference on Speech Science & Technology*, ed. P. Warren and C. Watson, 421–426. University of Auckland.

Szakay, Anita. 2007. Identifying Māori English and Pākehā English from Suprasegmental Cues: A Study Based on Speech Resynthesis. Master's thesis, University of Canterbury.

Thomas, Erik R., and Jeffrey Reaser. 2002. Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. *Journal of Sociolinguistics* 8:54–87.

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

Department of Linguistics
University of British Columbia
Totem Field Studios
Vancouver, British Columbia, V6T 1Z4
szakay@interchange.ubc.ca