

BLS 34, No 1 2008. DOI: <http://dx.doi.org/10.3765/bls.v34i1.3580>
(published by the Berkeley Linguistics Society and the Linguistic Society of America)

The Relative Importance of Rhythm and Intonation for the Perception of New Zealand English Dialects

ANITA SZAKAY

University of British Columbia, Canada

0. Introduction

The present paper is part of a larger research project that set out to investigate the production and perception of suprasegmental features of the two main ethnolects of New Zealand English. Maori English (ME) is spoken by the indigenous population, while the variety used mainly by people of European descent is referred to as Pakeha English (PE).

Previous research has demonstrated that these two ethnic dialects display distinct rhythmic qualities as well as differing intonational patterns. ME has been shown to be significantly more syllable-timed than PE (e.g., Warren 1998). Using the vocalic Pairwise Variability Index (PVI) (Grabe and Low 2002), Szakay (2006) also confirmed that ME (mean PVI = 46.4) is significantly less stress-timed than PE (mean PVI = 57.2), where a lower PVI indicates a more syllable-timed language and a higher PVI indicates a more stress-timed language.

The two varieties also differ in their use of the High Rising Terminal contour (HRT), which is a salient rise in pitch at the end of non-interrogative intonational phrases (uptalk). HRT has been extensively studied in New Zealand English, and is reported to be used mainly by young, female speakers. A major study carried out by Britain (1992) showed evidence that this pattern is used in different proportions by Maori and Pakeha. His analysis indicated that Maori speakers use a significantly higher percentage of HRTs than Pakeha. The results also revealed that young Maori men use levels similar to women, while young Pakeha men are extremely low users of HRTs.

The present study aimed to investigate whether listeners are aware of these prosodic differences and whether they are capable of tuning in to speaker rhythm and intonation to facilitate dialect identification. The study also hoped to determine the relative importance of the rhythmic and intonational patterns of a speaker for ethnic dialect categorization in New Zealand by establishing whether listeners rely more on rhythm or HRTs to identify a speaker's ethnicity.

Perception of Rhythm and Intonation in NZ English

2. Results

Logistic regression models were fit for each condition to reveal what suprasegmental features significantly affect listeners' perception of a speaker's ethnicity. The analysis was carried out using the R Statistical Package.

2.1. Rhythm

As reported in Szakay (2008b), rhythm was shown to be a significant predictor of perceived ethnicity in all conditions where it was retained in the speech signal. Syllable-timed speakers are identified as Maori sounding, while stress-timed speakers are more likely to be classified as Pakeha, indicating that overall New Zealanders are in fact aware of the rhythmic difference between Maori English and Pakeha English. The social network of a listener was also shown to be of importance for their ability to rely on rhythm. Listeners who are highly integrated into Maori social networks did consistently better at the dialect identification task than those who are not integrated (for more detail see Szakay (2008b)).

2.2. Intonation

The results regarding the production of HRTs confirm previous research demonstrating that, overall, Maori speakers use a significantly higher percentage of HRTs than Pakeha speakers (linear regression, $p < .05$).

The percentage of HRTs used by a speaker was also a significant predictor of perceived ethnicity in all conditions of the perception experiment where intonation was retained in the speech signal. Results for each of the four conditions are reported below.

2.2.1. Condition Seven – Normal Speech

In the unaltered speech condition, the regression model reveals that perceived ethnicity is affected by a significant interaction, which occurs between the percentage of HRTs and the speaker's actual ethnicity. This is shown in (1). The y-axis shows the log odds of a speaker being identified as Pakeha. A higher value indicates that the speaker is more likely to be perceived as Pakeha, while a low value indicates that the speaker is more likely to be identified as Maori. The coefficient table for the model of perceived ethnicity is shown in (2), while the corresponding anova table is presented in (3).

The percentage of HRTs seems to be a very important cue for dialect identification in the case of Maori speakers. If a Maori speaker has a low percentage of HRTs, then they are more likely to be identified as Pakeha. However, if they use a lot of HRTs then they are considered to be Maori sounding. Listeners do not make use of the percentage of HRTs in the case of Pakeha speakers. As the relatively flat line around zero indicates on the graph, the percentage of HRTs used by a Pakeha speaker does not seem to influence participants' responses in either direction. This presumably indicates that HRT use is a secondary cue in this condition, unlikely to influence responses alone, but increasing the certainty of the listener if a speaker has other characteristics that make them sound Maori.

Anita Szakay

1. Method

The methodology followed in this study is nearly identical to the one described in Szakay (2008b). Accordingly, 36 New Zealanders were used for the analysis of rhythm and intonation. The 24 Maori and 12 Pakeha speakers were recorded reading a passage as well as narrating a story.

The rhythmic patterns of the speakers were analyzed using the normalized vocalic PVI (Grabe and Low 2002). The total number of vocalic segments taken into consideration for the analysis of rhythm was 3281, based on both reading passages and narratives.

HRTs were analyzed in the spontaneous speech segments only, as speakers did not produce a great number of HRTs in the reading passages. HRTs were manually labeled on a separate textgrid tier in Praat, from the valley to the peak of the phrase final pitch rise. Both actual and potential occurrences were coded for, allowing for the computation of the percentage of HRTs used by each speaker. Of the 36 narrative passages, 16 did not contain any HRTs and were excluded from the analysis.

Twenty narratives (10 Maori, 10 Pakeha) were chosen to be included in the dialect identification experiment as stimulus material. Seven different speech conditions were created that all retained different suprasegmental cues in the speech signal. The present study concentrates on 4 of these conditions, that preserved the intonational patterns of the original passage:

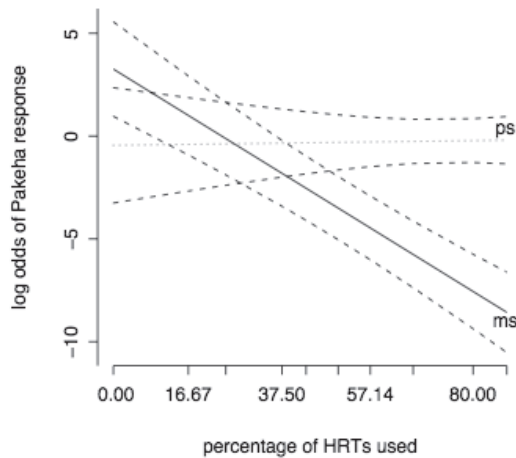
- (a) Condition Seven: Normal speech
- (b) Condition Six: Low-pass filtered speech
- (c) Condition Five: Resynthesized Rhythm and Intonation together
- (d) Condition Four: Intonation only

All conditions were created using the Praat acoustic analysis software (Boersma 2006). Low-pass filtering was carried out at 400Hz, thus retaining both the rhythmic and the intonational patterns of the speaker (and arguably voice quality features as well). Condition Five was created by first humming the signal, then replacing all consonants by silence. The intonation only condition also made use of Praat's humming option but only original pauses were replaced by silence. A more detailed description of how these conditions were created can be found in Szakay (2008a, 2008b).

55 Pakeha and 52 Maori listeners were asked to perform a forced-choice dialect identification task in all seven conditions. The results reported below concentrate on the above-mentioned 4 conditions.

Anita Szakay

- (1) Condition Seven – Model effects showing likelihood of Pakeha perceived ethnicity by percentage of HRTs used and actual speaker ethnicity.



- (2) Condition Seven - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-31.75	7.53	-4.21	0.0000
PVI	0.04	0.04	1.12	0.2607
PMII	-0.58	0.18	-3.24	0.0012
log SD of pitch	27.41	5.61	4.88	0.0000
speaker gender=male	32.50	7.57	4.29	0.0000
% of HRTs	-0.13	0.01	-7.95	0.0000
speaker ethnicity=Pākehā	-3.70	1.14	-3.25	0.0011
minimum pitch	0.02	0.00	3.21	0.0013
mean slope of HRTs	-0.01	0.00	-6.28	0.0000
PVI * PMII	0.01	0.00	3.10	0.0019
log SD of pitch * speaker gender=male	-24.71	5.38	-4.59	0.0000
% of HRTs * speaker ethnicity=Pākehā	0.13	0.02	6.08	0.0000

Perception of Rhythm and Intonation in NZ English

(3) Condition Seven - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
PVI	15.53	2	0.0004
PMII	11.51	2	0.0032
log SD of pitch	26.25	2	<.0001
speaker gender	24.26	2	<.0001
% of HRTs	65.21	2	<.0001
speaker ethnicity	64.82	2	<.0001
minimum pitch	10.31	1	0.0013
mean slope of HRTs	39.46	1	<.0001
PVI * PMII	9.64	1	0.0019
log SD of pitch * speaker gender	21.03	1	<.0001
% of HRTs * speaker ethnicity	36.97	1	<.0001
TOTAL INTERACTION	57.81	3	<.0001
TOTAL	385.18	11	<.0001

2.2.2. Condition Six – Low-pass Filtered Speech

As pitch movement is retained by low-pass filtering, HRTs are easily perceivable in this condition. However, not all participants seem to use the percentage of HRTs present to identify speaker ethnicity. Whether a speaker uses many HRTs or none at all does not seem to have an effect on the responses of Maori participants in this condition. This is illustrated in the left panel of (4) by the flat lines for both Maori male and female participants. The behavior of Pakeha males parallels that of Maori participants, as they do not seem to use HRTs as a cue either. Pakeha females, however, do tune in to the percentage of HRTs used by a speaker in this condition. For them a speaker who uses many HRTs sounds very Maori, while the lack of HRTs indicates that the speaker is more likely to be Pakeha. See the right panel of (4) for this gender difference. Coefficient and anova tables for the model in Condition Six are shown in (5) and (6) respectively.

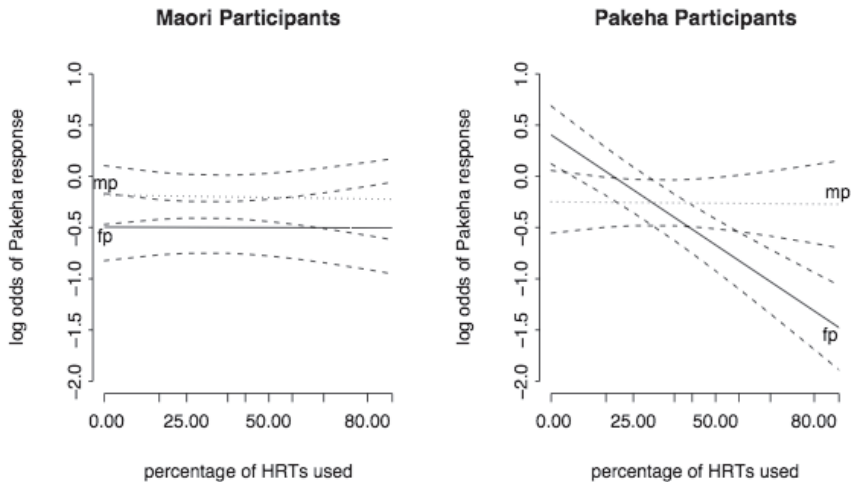
2.2.3. Condition Four – Intonation Only

HRTs were only used by Maori participants as a cue in this condition. For Maori listeners, the more HRTs a speaker uses, the more likely the speaker is to be identified as Maori sounding. Pakeha participants, on the other hand, do not make use of HRTs as a cue for dialect identification in the intonation alone condition: whether a speaker uses many HRTs or not does not influence Pakeha participants' responses in either direction. This is illustrated in (7) by the flat line marking

Anita Szakay

Pakeha responses. The corresponding coefficient and anova tables are presented in (8) and (9), that also reveal that Pakeha listeners tune into speakers' standard deviation of pitch instead of HRTs in this condition.

- (4) Condition Six – Model effects showing likelihood of Pakeha perceived ethnicity by percentage of HRTs used and listener gender. Coefficients set to Maori listeners in left panel, to Pakeha listeners in right panel.



- (5) Condition Six - Coefficient table for model of perceived ethnicity.

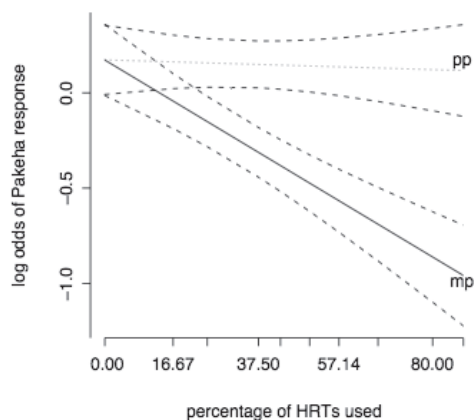
Variable	Coef	S.E.	Wald Z	P
Intercept	-5.42	0.85	-6.36	0.0000
% of HRTs	-0.00	0.00	-0.03	0.9759
participant ethnicity=Pākehā	0.89	0.24	3.74	0.0002
participant gender=male	0.31	0.19	1.58	0.1142
PVI	0.04	0.01	4.05	0.0001
PMII	-0.28	0.06	-4.57	0.0000
minimum pitch	0.01	0.00	6.31	0.0000
speech rate	0.52	0.11	4.44	0.0000
number of pauses	-0.12	0.03	-3.91	0.0001
% of HRTs * participant ethnicity=Pākehā	-0.02	0.00	-4.93	0.0000
% of HRTs * participant gender=male	-0.00	0.00	-0.09	0.9322
participant ethnicity=Pākehā * part gender=male	-0.96	0.27	-3.48	0.0005
PVI * PMII	0.00	0.00	4.50	0.0000
% of HRTs * part ethn=Pākehā * part gender=male	0.02	0.00	3.53	0.0004

Perception of Rhythm and Intonation in NZ English

- (6) Condition Six - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
% of HRTs	50.38	4	<.0001
participant ethnicity	26.43	4	<.0001
participant gender	28.90	4	<.0001
PVI	142.74	2	<.0001
PMII	20.92	2	<.0001
minimum pitch	39.76	1	<.0001
speech rate	19.74	1	<.0001
number of pauses	15.31	1	1e-04
% of HRTs * participant ethnicity	24.29	2	<.0001
% of HRTs * participant gender	24.13	2	<.0001
participant ethnicity * participant gender	14.35	2	8e-04
PVI * PMII	20.27	1	<.0001
% of HRTs * part ethnicity * part gender	12.48	1	4e-04
TOTAL INTERACTION	59.63	5	<.0001
TOTAL	279.5	13	<.0001

- (7) Condition Four – Model effects showing likelihood of Pakeha perceived ethnicity by percentage of HRTs used and listener ethnicity.



Anita Szakay

(8) Condition Four - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-1.40	0.25	-5.52	0.0000
log SD of pitch	1.13	0.19	5.92	0.0000
participant ethnicity=Pākehā	0.00	0.12	0.01	0.9948
% of HRTs	-0.01	0.00	-6.06	0.0000
participant ethnicity=Pākehā * % of HRTs	0.01	0.00	4.32	0.0000

(9) Condition Four - Anova table for model of perceived ethnicity.

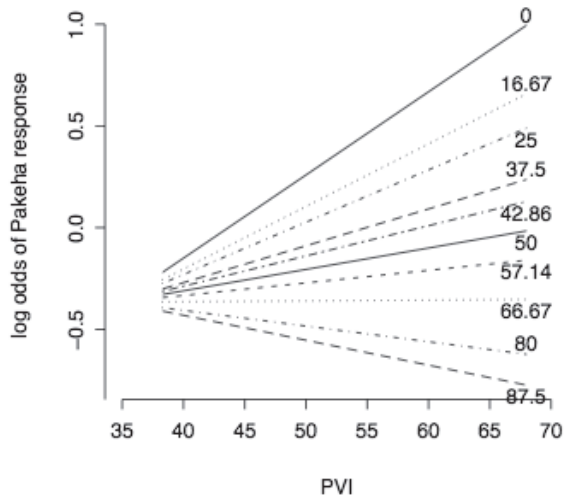
Factor	Chi-Square	d.f.	P
log SD of pitch	35.02	1	<.0001
participant ethnicity	38.14	2	<.0001
% of HRTs	36.67	2	<.0001
participant ethnicity * % of HRTs	18.67	1	<.0001
TOTAL	78.15	4	<.0001

2.3. Interaction Between Rhythm and Intonation

Condition Five reveals an interaction between rhythm and the percentage of HRTs used by a speaker. If a speaker is very syllable-timed, they are identified as Maori. In this case, it does not matter what percentage of HRTs the speakers uses. However, if the speaker is more stress-timed, then the perceived ethnicity is in fact determined by the frequency of HRTs. This operates in the following way. If the stress-timed speaker does not use HRTs, they will be identified as very Pakeha. However, if the stress-timed speaker does use a lot of HRTs, they will be perceived as Maori, despite having a high PVI value. That is, the frequency of HRTs seems to be more important a cue than rhythm itself, *if* the speaker is stress-timed. On the other hand, rhythm is more important than HRTs *if* the speaker is very syllable-timed. The interaction between rhythm and the percentage of HRT is shown in (10). Coefficient and anova tables for this model are presented in (11) and (12).

Perception of Rhythm and Intonation in NZ English

- (10) Condition Four – Model effects showing likelihood of Pakeha perceived ethnicity by PVI and percentage of HRTs used.



- (11) Condition Five - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-1.18	0.67	-1.77	0.0774
PMII	-0.21	0.05	-3.83	0.0001
PVI	0.01	0.01	0.98	0.3275
log SD of pitch	0.75	0.19	3.90	0.0001
% of HRTs	0.02	0.01	1.84	0.0654
PMII * PVI	0.00	0.00	3.65	0.0003
PVI * % of HRTs	-0.00	0.00	-2.84	0.0045

Anita Szakay

(12) Condition Five - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
PMII	15.58	2	0.0004
PVI	34.39	3	<.0001
log SD of pitch	15.24	1	0.0001
% of HRTs	62.93	2	<.0001
PMII * PVI	13.31	1	0.0003
PVI * % of HRTs	8.06	1	0.0045
TOTAL INTERACTION	21.09	2	<.0001
TOTAL	97.20	6	<.0001

3. Discussion

As discussed in the Szakay (2008b), rhythm seems to be a reliable cue for ethnic dialect identification in the New Zealand context. However, the use of HRTs is just as useful a cue for the listener. Previous research has shown that ME speakers are likely to use a higher percentage of HRTs than Pakeha speakers (e.g., Britain 1992). The results from the production experiment in this study also indicated that overall Maori speakers use significantly more HRTs. The results also reveal that listeners' perception is in line with the findings of these studies, in that participants generally expect Maori speakers to use more HRTs in their speech than do Pakeha speakers. The percentage of HRTs used by the speaker is a variable that showed a significant effect in all conditions where intonation was retained in the speech signal.

In Condition Seven, where all information, including segmental features, was preserved in the speech signal, the percentage of HRTs interacting with speaker ethnicity had a significant effect on perceived ethnicity. Only the percentage of HRTs used by Maori speakers influenced participants' responses. The more HRTs a Maori speaker uses, the more Maori sounding they will be marked by the listener. However, participants do not make use of HRTs in the case of Pakeha speakers. This suggests that the segmental features of Pakeha English are more important for ethnic dialect identification and as such override the frequency of the HRTs used by the speaker.

In Condition Six, low-pass filtered condition, where segmental information is no longer available, only Pakeha females are tuning into the percentage of HRTs. The direction of their responses is the same as in all other conditions, that is the more HRTs a speaker uses, the more Maori sounding they are perceived to be. Pakeha men and Maori participants are tuning in to cues other than HRTs when the low-pass filter is applied (e.g., rhythm, speech rate, minimum pitch and the number of pauses) (see Szakay 2008a).

Condition Five demonstrates that HRTs in some instances serve as more useful an indicator of perceived ethnicity than rhythm itself. In this condition a stressed speaker's perceived ethnicity will depend on how often they use HRTs.

Perception of Rhythm and Intonation in NZ English

Again, more HRTs are an indicator of a more Maori sounding speaker. In the case of syllable-timed speakers, however, rhythm is more important a factor for listeners than the percentage of HRTs. In Condition Five the statistical model predicts that all participants use these variables the same way, regardless of their ethnicity. In contrast, only Maori participants rely on HRTs in the intonation only Condition Four. Again, a greater number of HRTs suggests that the speaker is more Maori sounding. Pakeha listeners do not make use of HRTs in this condition to distinguish between the two dialects. Instead, they rely on the speaker's standard deviation of pitch.

The results discussed in Szakay (2008b) also revealed that participants who have had more exposure to ME are more capable of relying on syllabic rhythm in a dialect identification task, even in the lower conditions, where those participants with less exposure seem to be tuning into speech rate instead of rhythm itself. These results prove that different listeners utilize different cues for ethnic dialect identification. Thomas and Reaser (2002) argue that it is also possible that different cues are used to identify different speakers. The results from the present study support this claim showing that speakers who are very syllable-timed will be identified as Maori based on their rhythmic characteristics. Those speakers who demonstrate more of a stress-timed rhythmic pattern, however, will be instead judged on their use of HRTs. They might still be identified as Maori regardless of being stress-timed, if they use a high percentage of HRTs.

4. Summary

Rhythm and the percentage of HRTs used by the speaker proved to be extremely useful cues for participants in the dialect identification task. In the case of these two prosodic features, not only do we have production evidence of a divergence between ME and PE, this evidence is supported by perception evidence that shows that listeners are aware of these features and associate them with ethnicity. The acoustic analysis in Szakay (2006) demonstrated that ME speakers are significantly more syllable-timed than PE speakers, while the dialect identification experiment in Szakay (2008b) showed that listeners are aware of this difference and are able to rely on rhythm in ethnic dialect differentiation.

The production data also confirmed the results of previous research on the use of HRTs in New Zealand (e.g., Britain 1992), showing that, overall, Maori use a significantly higher percentage of HRTs than Pakeha. Listeners are aware of this tendency and expect a high number of HRTs to be an indicator of a Maori speaker. Moreover, the perception results also revealed the relative importance of rhythm and HRTs. Listeners heavily rely on rhythm in their ethnic judgments when the speaker is very syllable-timed. These speakers will be identified as Maori, regardless of their use of HRTs. However, if the speaker is stress-timed, it is in fact the frequency of HRTs that will be the deciding factor in distinguishing ME from PE. Thus, we can conclude that both of these prosodic features facilitate accurate dialect identification in the New Zealand context.

Anita Szakay

References

- Boersma, Paul, and David Weenink. 2006. Praat: Doing Phonetics by Computer. Version 4.4.12. Computer Program, <http://www.praat.org/> (accessed March 1, 2006).
- Britain, David. 1992. Linguistic Change in Intonation: The Use of High Rising Terminals in New Zealand English. *Language Variation and Change* 4(1):77-104.
- Grabe, Esther, and Ee L. Low. 2002. Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven and N. Warner, eds., *Laboratory Phonology 7*, 515-546. Berlin: Mouton de Gruyter.
- Szakay, Anita. 2006. Rhythm and Pitch as Markers of Ethnicity in New Zealand English. In P. Warren and C. Watson, eds., *Proceedings of the 11th Australasian International Conference on Speech Science & Technology*, 421-426. Auckland, New Zealand: University of Auckland.
- Szakay, Anita. 2008a. *Ethnic Dialect Identification in New Zealand: The Role of Prosodic Cues*. Saarbrücken, Germany: VDM Verlag Dr. Müller.
- Szakay, Anita. 2008b. Social Networks and the Perceptual Relevance of Rhythm in Maori English: A New Zealand Case Study. In K. Gorman, ed., *U. Penn Working Papers in Linguistics 14.2: Papers from New Ways of Analyzing Variation 36*, 148-156.
- Thomas, Eric, and Jeffrey Reaser. 2002. Delimiting Perceptual Cues used for the Ethnic Labeling of African American and European American Voices. *Journal of Sociolinguistics* 8(1):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.

Anita Szakay
University of British Columbia
Department of Linguistics
Totem Field Studios
Vancouver, BC, V6T 1Z4

szakay@interchange.ubc.ca