DEVELOPING A METHODOLOGY FOR ELICITING AND ANALYZING PERCEIVED PROSODIC PROMINENCES: A MĀORI CASE STUDY

Laura Thompson, Catherine I. Watson, Helen Charters

This paper outlines a methodology developed for eliciting and analyzing perceived prosodic prominences, based on a study of Māori, the indigenous language of New Zealand. Māori has some existing description of expected stress patterns and pitch contours, including which syllables may be expected to be prominent (e.g. Biggs 1969, Bauer 1993). Owing partly to the influence of English, there exists the perception that the sound of the language has changed over time. This has been observed by listeners of all proficiency levels, from fluent speakers to first-timers, and is found in the recordings from different time periods collected in the MAONZE database (King et al. 2011).

The principal aim of the study described here was to determine where listeners hear prosodic prominences in Māori, and examine what it is acoustically that makes them hear certain units (in this case, syllables) as prominent.

In the study, a set of 30 utterances from three different time periods were heard by 92 listeners. The utterances were continuous speech, taken from interviews in the MAONZE database: 10 each from male Historical Elders (b. 1880s, rec. 1940s), male Present Day Elders (b. 1920-30, rec. 2000s), and male Young L1 speakers (b 1970-90, rec. 2000s). The study participants were divided into three roughly equal-sized groups according to their self-rated Māori proficiency: ‘High’ (28) meant conversational ability from fluent to basic; ‘Exposed’ (30) meant word-and-phrase or parsing level ability and was the category for most New Zealanders. ‘Zero’ (34) meant no or low exposure, and applied to most of the overseas participants. All participants had high English proficiency. They were sourced via advertising, networks, and social media applications. As highly proficient speakers of some languages (including Māori) are difficult to come by, sourcing participants is a key problem in a study of this type; the others being selection of stimuli, appropriateness of delivery method, and result-wrangling.

First, the stimuli. Here, they had to be chosen based on clarity, completeness, length and content. Though 'perfect' delivery is very hard to obtain, interruptions in the source recordings will affect an inherently relative concept such as prominence. To find a balance between participant stamina and the opportunity to test enough tokens, study size was set at 32 utterances. They are a mixture of simple and complex sentences. In hindsight, most participants could probably have handled twice the number of stimuli.

The delivery method was a web form interface. The utterances were presented as shown in Fig. 1, with a checkbox for each syllable. Each utterance could be played as many times as the participant needed, but not paused and restarted. Participants were simply asked to listen and check boxes for the syllables that they thought ‘stood out’. The whole survey took most participants about half an hour.

We chose this method for two main reasons: to increase efficiency and distributability of the study, and to take pressure off the participants. An earlier pilot study, employing a recorded

![Fig. 1. Partial screenshot from the web survey.](image-url)
tapping exercise (Thompson 2009), required more equipment, time, and the physical presence of both the participant and the researcher. It induced performance anxiety in some participants, who felt that the method measured physical coordination rather than intention, and a delay was apparent. The web survey let participants operate in their own time, without supervision, and was globally accessible, allowing far more participants. Behind the scenes, raw result processing is also much more efficient, with binary results connected directly to specific units within the utterance.

All syllables that received more votes than those adjacent to them are considered to be ‘perceived prominences’, but this list is then reduced statistically to only those which have a significantly higher number of results than their neighbors. The final set of significant perceived prominences (or, ‘PPMs’) is then further examined in several ways. Firstly, a durational analysis of ‘intervals and events’ (meaning PPMs and the time between them) looks for patterns of prominence that may create ‘rhythm’. Secondly, the location of PPM syllables within the utterances is compared with the predicted stress locations in those utterances, to see how closely they match. Thirdly, there is acoustic analysis of duration, pitch, vowel quality and loudness, in which the PPM syllables are assessed for any exceptional behavior in these areas. Additionally, we look for patterns and differences in response to the speech from the older versus the more recent speakers, and between the responses of the three participant proficiency groups (native versus non-proficient).

One discovery of the study has been that even though native speaker (or in this case, ‘high’ proficiency) perception is important for the analysis of prosody, these speakers do not always provide the clearest results from an acoustic point of view. For Māori, at least, the higher the proficiency, the less likely the listener is to find specific syllables prominent. Overall, ‘high’ proficiency speakers checked boxes more often than the other groups, which, perversely, led to a smaller number of PPMs recorded for that group. This is probably due to their understanding the meaning: they are less tuned to the sound than to the content, so when asked to pick what ‘stands out’, they are less selective. The impact of the acoustic cues is not as strong as for lower proficiency groups, who have only what the sound gives them. Both the influence of English on Māori and the fact that all the listeners spoke good English mean that they may all be tuned to the same cues. Only when semantic content is a factor do they focus on them less. A listener can't be forced to focus on sound without the use of filtered speech, and the web survey format presents difficulties here. They are not insurmountable, but were out of the scope of the current study.

The methodology remains under development, even while producing results for the current study. The ultimate goal is a description or illustration of prosodic prominences in a given language, and the methodology intended to be modifiable and applicable to any language that could benefit from such description.

References