Acoustic and perceptual correlates of the Hiaki pitch accent

Florian Hafner

The University of Arizona
Department of Linguistics
P.O. Box 210025
Tucson, AZ, USA

florianhafner@email.arizona.edu

Abstract

This study examines the acoustic and perceptual correlates of the pitch accent of the undocumented and endangered language Hiaki with two different approaches: acoustic measurements and a perceptual study. Fifty-five tokens were elicited from one native Hiaki speaker. Measurements were made in Praat for vowel length, pitch peak, intensity peak, and a potential alignment of a high tone in terms of the offset of the first vowel and as function of vowel length. While *naamu*-cases showed significant difference in pitch peak, intensity peak, and vowel length, it was particularly the *namuta*-cases which exhibited no difference between the vowels. Therefore, 36 native English speakers were asked to mark prominent syllables in individual Hiaki words and sentences to avoid lexical biases and native speaker influences. This perception experiment showed that the long vowels in *naamu*-cases serve as massive prominence anchor but listeners were unsure were to mark prominence in *namuta*-cases. Consequently, the combination of both experiments could not identify and locate precise acoustic correlates and the placement of the morphologically determined pitch accent in Hiaki so that further research is necessary to provide a better understanding.

Keywords: *Hiaki*, *pitch accent*, *perception, prominence, LMEDS, rapid prosodic transcription*

1. Introduction

Languages around the world vary considerably in how they create prosody and mark prominence. Speakers can use duration, intensity, pitch, or a combination of thereof to convey messages, highlight syllables, or mark the rhythm of their language. This paper examines the acoustic and perceptual correlates of prominence and its placement for the severely undocumented and endangered language Hiaki. While Hiaki possesses both, a lexical pitch accent and a morphologically determined pitch accent, this study will only analyze the morphological one.

1.1 Overview of Hiaki

Hiaki, Yaqui, Yoeme, or Yoem Noki, is a Uto-Aztecan language, branching into Southern Uto-Aztecan, Taracahitic, and Cahitan making Mayo its closest relative. Mayo and Hiaki are mutually intelligible. Their close relation can be found in the earliest printed description of Hiaki in 1737; it was published by Francisco Javier and compares "remarkably well [with] contemporary Mayo" (Dedrik & Casad, 1999, p. 3). Mayo has a pitch accent system which will be outlined later as it might provide insights in how the pitch accent in Hiaki could work. Hiaki is spoken primarily in the Mexican state of Sonora, along the Yaqui River, and in the US state of Arizona, particularly around Tucson. It is an endangered language with about 100-500 speakers in Arizona and approximately 17,000-21,000 in Sonora. Ethnologue classifies it as 8a – moribund (Simons &

Fennig, 2017). Most speakers are bilingual in Hiaki and Spanish and are often not literate in the two. It is an SOV language with a relatively small phoneme inventory; the system comprises five vowels, /i, e, a, o, u/, as well as 15 consonants. The phonotactics are similar to Spanish in that word-initially, the language allows CV- and CVC-, and -V- and -VC- word-medially and -finally (Dedrik & Casad, 1999, p. 28). With regards to morphophonological processes, Dedrik and Casad (1999, p. 28-30) list echo vowels, pitch shift, the dropping of intervocalic /l/ and /r/, palatalization, velarization of /s/, glide reduction, and reduplication.

1.2 The Hiaki pitch accent

Hiaki possesses two different types of pitch accents: one is a lexically determined pitch accent and the other is a morphologically determined pitch accent. The following two sections will outline how those pitch accents are realized, whether acoustics correlates are known, and where the pitch accents attach to.

1.2.1 The lexically determined pitch accent

Hiaki possesses a lexical pitch accent creating minimal pairs such as *héeka* ‘wind’ and *heéka* ‘drank’. Demers et al. (1999, p. 41) state that such pairs are true minimal pairs and phonological conditioning can be ruled out as the sole difference between the words is the pitch accent. This implies that the pitch accent is lexically specified and not conditioned by anything else, e.g. the morphological structure of the word. While acoustic measurements illustrate a distinct pitch accent, i.e. a clear high tone on the accented mora, only a few of those minimal pairs are attested. Fabián and Duarte (2006, p. 10) list the ones found in table 1. Researchers often additionally mention *wáate* 'remember' and *waáte* 'people' as another minimal pair but it turned out that this pair is not a minimal pair for our consultant because *waáte* 'people' is pronounced as [wate] with the first vowel being short.

Table 1: This table shows the minimal pairs for the lexical pitch accent

yóoko	'stripped'[1]	*yoóko*	'tomorrow'
téeka	'sky'	*teéka*	'to lay it flat'
káate	'build a house'	*kaáte*	'walk'
ánia	'world'	*anía*	'help'

The minimal pairs above demonstrate that the lexical pitch accent attaches to morae. In *káate* 'build a house', the pitch accent goes on the first mora whereas it attaches to the second mora in *kaáte* 'walk'.

In order to confirm this lexical pitch accent and the fact the consultant reported later reliably produces it, several of the minimal pairs above were elicited, for instance, *yóoko* 'leopard' and *yoóko* 'tomorrow'. *Wáate* 'remember' and *waáte* 'people' is not a minimal pair for the

[1] *Yóoko* also means 'leopard' or 'jaguar'

consultant because *waáte* ‘people’ is produced with a short first vowel, as in [wate] and not [wa:te]. The elicitation session provided *ánia* ‘world’ and *anía* ‘help’. The pitch on *ánia* ‘world’ was almost a flat line in both elicited tokens. It is clearly different from *anía* ‘help’ with a pitch peak of 167Hz on the first vowel [a] and 192Hz on the second vowel [i].

Another example is *yóoko* ‘leopard’ and *yoóko* ‘tomorrow’. Figure 1a and figure 1b illustrate the two words. One can see a slight peak on the first mora of *yóoko* ‘leopard’ in figure 1a. The pitch then drops to 150Hz at the end of the second mora before the [k]. The pitch on the

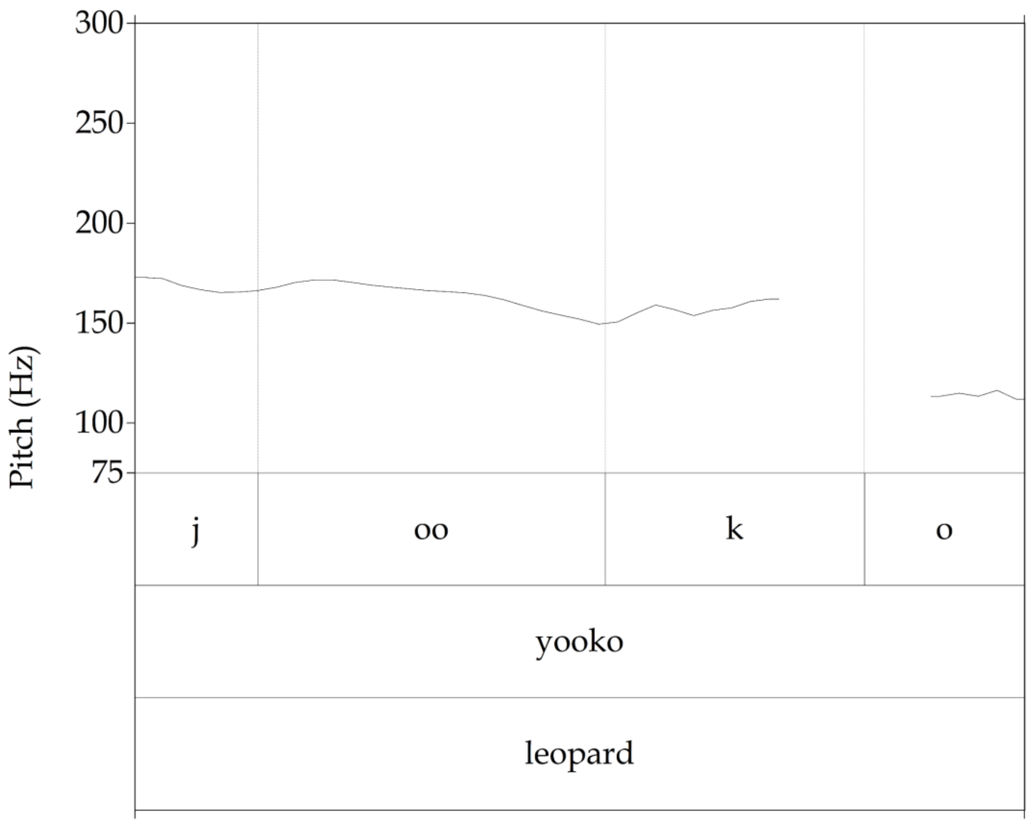

Figure 1a: Pitch track of *yóoko* ‘leopard’ demonstrating a pitch peak on the first mora [o] with a following fall over the second mora. The pitch on the third mora, the last segment [o], is clearly lower as well.

third mora, which is [o], peaks at 114Hz. Comparing this to the second example *yoóko* ‘tomorrow’ in figure 1b, the pattern looks strikingly different. A steep pitch rise characterizes the first two morae, i.e. the long first vowel [o:], peaking at 260Hz; this is a 61Hz rise from the offset of the first segment /j/. One can also observe a lower pitch on the third mora with a clear fall starting at 211Hz. Consequently, the consultant reliably produces the lexical pitch accent.

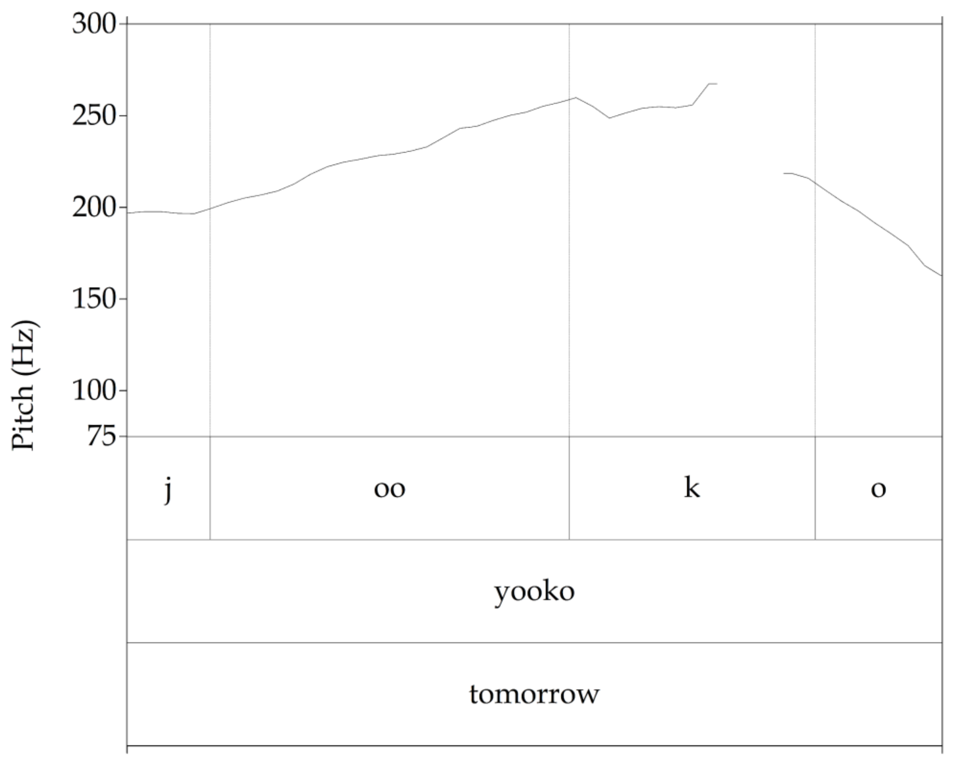

Figure 1b: Pitch track of *yoóko* ‘tomorrow’. Visible is the lexical pitch accent peaking at 260Hz on the second mora.

1.2.2 The morphologically determined pitch accent

Hiaki additionally possesses a morphologically determined pitch accent. The pitch accent generally occurs when a word containing a long vowel is affixed. For instance, adding the future tense suffix *-ne* to *waawa* ‘roast in coals’ shortens the long vowel to *wawane* ‘will roast in coals’. The question is then: how does the pitch accent behave? As Demers et al. (1999, p. 46) claim, the pitch accent may shift so that its overall position in the word remains the same. Those cases then provide a perfect environment to examine whether the pitch accent attaches to morae or syllables. Admittedly, one could argue that the morphologically determined pitch accent simply behaves as the lexically determined pitch accent and attaches to morae[2]. This would imply that under affixation, which shortens the long first vowel, the pitch accent would jump to the second syllables of an affixed word form because the second mora will then be in the second syllable (compare *waawa* ‘roast in coals’ versus *wawane* ‘will roast in coals’, the second mora being underlined).

[2] Pitch accent languages can generally vary whether they will realize a pitch accent with regards to morae or syllables. Tokyo Japanese is one variety that attaches its pitch accent to morae (Cutler & Otake, 1999). Serbio-Croatian, a stress and pitch-accent language, behaves similarly and attaches its pitch accent to morae as well (Inkelas & Zec, 1988). Swedish, on the other hand, aligns its two pitch accents, grave and acute, in relation to stressed syllables (Bruce, 2005). It is therefore not unlikely to see a language such a Hiaki to attach the morphologically determined pitch accent to either morae or syllables.

Figure 2 outlines the difference between a morphologically determined Hiaki pitch accent attaching to either morae or syllables. If the pitch accent attaches to the second mora, it would

Mora-Account		**Syllable-Account**	
muúmu	*mumú-ta*	muu.mú	mu.mú.-ta
μμ μ	μ μ μ	σ σ	σ σ σ
hive	hive-acc	hive	hive-acc
'hive'	'hive'	'hive'	'hive'

Figure 2: The left shows a mora-account assuming the pitch accent attaches to the second mora. It would remain on the second mora under affixation, shifting to the second syllable. By contrast, this shift would be absent if the pitch accent attached to the second syllable.

shift to the second syllable under affixation as the shortening of the first vowel causes the second mora to occur in the second syllable, as in *mumúta* 'hive-acc'. On the other hand, if the pitch accent attaches to the second syllable in first place, no such shift would be expected as the second syllable remains the second one throughout affixation processes. In case the pitch accent would attach to either the first mora or syllable, no shift would be expected either. Until now, no study has found a clear answer to this question and most previous research could not determine the precise acoustic correlates (Fabián & Duarte, 2006; Hagberg, 2008). Rather, studies often contradict in their findings.

1.3 Acoustic correlates and placement of the morphologically determined pitch accent

When looking at the previous research on Hiaki prosody, one will quickly realize that the terms *stress* and *pitch accent* are not used in a consistent matter, sometimes even interchangeably, which then complicates the matter when doing research in this area. This section will therefore look at English and Japanese first in order to lay a foundation of how stress and pitch accent are similar or different. It will then review past studies of the prosodic system in Hiaki highlighting the acoustic cues found so far and where Hiaki places its morphologically determined pitch accent.

Gussehoven (2004, p. 13-19) reviews English stress highlighting that a stressed element is not simply a louder pronunciation thereof. Stress is a feature of syllables that, in English, has an influence on vowel quality, duration, and intensity. Unstressed vowels are usually reduced, and stressed syllables commonly exhibit greater duration and intensity. However, the most important cue for English stress is the F0 and stressed syllables additionally serve as potential anchors for intonational tones.

Looking at a classic pitch accent language, Japanese, Haraguchi (1988) states that "strength and prominence is not a distinctive characteristic of pitch-accent in Japanese nor in any other pitch-accent languages" (p. 125). It is merely the F0 that cues a pitch accent. Haraguchi (1988) also suggests that even though the pitch-accented mora is longer, this difference is negligible as speakers are unable to perceive this lengthening. As a consequence, stress can be cued by a wide variety of factors such as duration, intensity, vowel reduction, pitch, or a complex interplay thereof whereas pitch accent is usually cued by the F0 only.

We will now return to the acoustic correlates and the placement of the Hiaki pitch accent. The first account about Hiaki prosody comes from Fraenkel (1959), who identifies prominence as stress rather than pitch accent[3]. He argues that stress is unpredictable but goes on the second first and second second syllable of a bisyllabic word while all monosyllabic words are stressed (Fraenkel, 1959, p. 8-11), for example [ʔoʔów] 'man' or [pípím] 'breast'. Trisyllabic words carry stress either on the first and third syllable or on the second only. The patterns would look like CV́CVCV́ for [kútaná] 'neck' or CVCV́CV [joémen] for 'men'. He also provides a detailed account of how Hiaki treats its vowels and consonants with regards to morae and syllables. Yet, more recent studies have not pursued his approach that Hiaki uses stress to cue prominence and Fraenkel does not mention anything about potential acoustic correlates. Crumrine (1961) also describes Hiaki prosody but only states that "stressed syllables are equally timed relative to each other regardless of the number of unstressed syllables between them" (Crumrine, 1961, p. 5). To

[3] Note here that when talking about stress or pitch accent in the following, different studies may define stress and pitch accent differently than outlined in the section about English stress and Japanese pitch accent. They may also have found different cues for either stress or pitch accent. However, each section reviewing one approach will outline what the researchers assumed to be stress or pitch accent and how it was cued.

conclude, this early research suggests that prominence, for which no acoustic correlates are known yet, attaches to syllables at the beginning of a word and that Hiaki is a stress-timed language.

Coming to more recent studies with more advanced phonetic methods, Fabián and Duarte (2006, p. 4) analyzed intensity and length, which they refer to as stress, and peak F0, which they refer to as pitch accent. Their conclusion is that pitch accent is phonological, i.e. contrastive as it can change the meaning of a word. They also state that stress (i.e. length but in particular intensity) is a non-contrastive, rhythmic device, suggesting that intensity and length serve to build appropriate foot structures for an ensuing pitch accent allocation. Note that this might seem unusual that a language has both lexical stress and a pitch accent but the authors argue in favor of Hiaki being a "lexical stress-accent language" (Fabián & Duarte, 2006, p. 1)[4]. The authors then propose that the unmarked Hiaki pattern on bi-syllabic words is a double aligned trochaic foot, i.e. strong weak [S W] on the foot level but aligned with a low high [L H] pitch accent. In terms of foot level, the left edge must contain a stressed syllable to make it a trochaic foot. By

[4] There are few other languages which also exhibit lexical stress and pitch accent at the same time, e.g. Swedish or Serbo-Croatian. Bruce (2005) distinguishes between rhythmical and intonational prominence in Swedish stating that stress belongs to the rhythmical part of intonation whereas the pitch accent belongs to the intonational patterns. Swedish stress is realized as a complex interplay between duration, spectrum, and intensity. The pitch accent, as part of a higher level of prominence (foot- or word-level), then amplifies certain acoustic aspects. Nevertheless, there are generally not many languages which make use of both lexical stress and pitch accent.

contrast, the right edge must contain a syllable with a high tone, i.e. [L H], which then makes both edges of a word prominent. The authors present a spectrogram of *he'e* 'drink' with a pitch peak of 151.5Hz on the second [e] (compared to 142.1Hz on the first [e]) but an intensity peak of 79.1db on the first [e], compared to 77.5db on the second [e][5]. The marked pattern, on the other hand, would be a single aligned trochaic or iambic foot, i.e. [S W] aligned with [H L] tones or in case of an iambic foot, [W S] aligned with [L H] tones. Overall, Fabián and Duarte's (2006) analysis shows that the Hiaki pitch accent is realized as a high tone that attaches to syllables. They also suggest that stress, manifested as length and intensity, is critical for the rhythmic structure of the language and the foot building processes.

Hagberg (2008) also investigated the morphologically determined pitch accent in Hiaki. His study agrees with Fabián and Duarte's (2006) in that the pitch accent he found also attaches to either the first or the second syllable, speaking for a syllable account when it comes to the pitch accent placement. In terms of acoustic correlates, Hagberg's research is contradictory to Fabián and Duarte (2006). Hagberg (2008) analyzed vowel length, F0, and intensity in interviews

[5] I would suggest to remain cautious regarding the intensity measures. The authors present spectrograms in which the intensity between accented and unaccented syllables varies considerably. Some instances show a peak difference of more than 8db whiles others exhibit less than 0.2db of a difference. It remains questionable whether such small peak differences are of any significance. As Bogomolets (2014) outlines, intensity may not be the best cue for prominence because the signal can be affected by many factors quite easily, e.g. the distance between the speaker and the microphone. Presenting a difference of 0.2db as evidence for a peak on a certain vowel may be questionable.

in Hiaki speakers from Sonora. His conclusion is that the F0 is the only reliable cue for the pitch accent; length and intensity do not play any role. He furthermore suggests that the pitch accent is marked by a high tone but he also observed a mid and a low tone; the mid tone occurs only on the first syllable of unaccented words whereas the low tone goes on the phrase-final syllable and does not spread to any other syllable. His ultimate verdict is that pitch accent, which has been called stress by other scholars, is cued by a rise in pitch occurring on the first or second syllable.

Next are Dedrick and Casad (1999)[6]. The authors' research also confirms the previous studies in that the Hiaki pitch accent attaches to syllables. They argue that Hiaki has two distinctive pitches, a high pitch and a low pitch; no precise acoustic correlates are mentioned but one may assume that the authors refer to the F0 as they are talking about high and low tones. The pitch then attaches to the first or second syllable. They furthermore make clear that in case the pitch accent goes on a long vowel, as they say a geminate, it can attach to either of the two vowels. However, they later claim that in case of affixation and the loss of the geminate vowel, the high pitch will shift to the following syllable (Dedrick & Casad, 1999, p. 29). This, then,

[6] The authors also included stress in their analysis claiming that stress is not prominent, i.e. it is not used by Hiaki speakers to indicate prominence. They list two exceptions though. One is emphasis and the other is a small number of words which contain a closed low-toned syllable, e.g. *tekipannoa* 'to work', where the pitch accent goes on [kí] but [ˈpan] is the stressed syllable.

would in fact support a mora-based explanation although the remaining parts of their writing seem to support a syllable account.

The paragraphs above have outlined potential acoustic correlates with no uniform results. One can surmise that in Dedrick and Casad (1999), it is likely the F0 that cues the pitch accent. Hagberg (2008) also concludes that the F0 is the sole reliable cue for the Hiaki pitch accent. This, however, stands in direct contrast to Fabián and Duarte (2006) who claim that the pitch accent is cued by the F0 but intensity and length are also crucial for forming appropriate foot structures. Regarding the placement of the pitch accent, all the studies have highlighted that the pitch accent is likely to attach to either the first or the second syllable. We will now turn to the question whether the pitch accent could also attach to morae instead of syllables.

Thinking of the few minimal pairs outlined earlier, e.g. *káate* 'build a house' and *kaáte* 'walk', the idea that the pitch accent may attach to morae is reasonable as well. As just outlined, Dedrick and Casad (1999, p. 29) observe that when affixation shortens the long first vowel, the pitch may shift. This would imply that the pitch accent attaches to morae rather than syllables. Demers et al. (1999) argue in the same vein. They likewise think that prominence is marked by a high tone, attributing little relevance to length and amplitude. Their data show that the amplitude across vowels remains relatively constant. The authors state that a "[h]igh tone is assigned to the first available vowel, provisionally assuming that long vowels are sequences of two identical short vowels" (Demers et al., 1999, p. 42-43). Each short vowel is then one mora, which is where the pitch accent attaches to. If the high tone goes on the second and not the first syllable, the authors consider the first vowel extrametrical. One observation that supports their claim is that more than 50% of the words "have initial high tone as second syllable tone" (Demers et al., 1999, p. 42). If the word receives an affix as demonstrated earlier in figure 2, the pitch accent will shift but simultaneously remain on the second mora. Interestingly, they add that this mora-loss rule which "shortens" the long vowel is necessary because they see vowel duration as a competing marker of prominence (Demers et al., 1999, p. 46); this might be of importance later as they suggest that length can be used to enhance prominence. Finally, they contend that prominence attached to morae in Uto-Aztecan languages is fairly common, mentioning above all Hiaki's closest relative Mayo. In general, Uto-Aztecan languages prefer prominence on the second mora which would perfectly fit into the mora-bases account.

More evidence in this case comes from Hagberg (1993), who thoroughly analyzed Mayo's prosodic system; he identifies it as a stress system. His analysis is in terms of stress as an autosegment that floats; autosegment means that stress "exist[s] on a plane by themselves without any link to individual segments" (p. 230). With the help of Mayo reduplication, (degenerate) foot structures and cyclic foot-building, derived vowel length, the minimal word requirement of two morae, as well as phrase-final extrametricality, he proposed that stress in

Mayo attaches to morae instead of syllables. Ultimately, he suggests that Mayo stress is manifested by a peak in the F0 track while he observes an abrupt pitch rise in the transition between a stressless and stressed syllable (Hagberg, 1993, p. 292).

This study will address two questions now: it will first examine the acoustic correlates of the pitch accent and then investigate where this pitch accent attaches to, i.e. to morae or syllables. In experiment one, data was elicited from a native Hiaki speaker and acoustic measurements were conducted in order to determine the correlates for the pitch accent in Hiaki. Earlier research judged the F0 to be the most reliable cue for the pitch accent (Demers et al., 1999; Hagberg, 2008). This experiment examined vowel length, pitch peak, and intensity peak in order to investigate a range of potential acoustic correlates. Regarding peak alignment, raw measurements from the offset of the first vowel to the pitch peak as well as pitch peak as a function of vowel length were also analyzed.

As these data remain inconclusive, the second experiment looked at the pitch accent from a perceptual side to supplement the acoustic data. It is not possible to recruit enough fluent Hiaki speakers to participate in a perception experiment, and the main questions are about the acoustic correlates and the pitch accent placement. Therefore, English-speaking listeners with no knowledge of Hiaki, and thus no lexical biases about prominence location, were used[7]. Thirty-six native English speakers were tested with the Rapid Prosodic Transcription method (RPT, e.g. Cole et al., 2016) to investigate where such speakers would locate prominences in isolated words and in short Hiaki prosodic phrases.

2. Experiment 1

The past literature has been inconclusive about the precise acoustic correlates of the pitch accent in Hiaki. While Hagberg (2008) found that F0 is the sole reliable cue, Fabián and Duarte (2006) included intensity and length arguing that this indicates stress, another important prosodic factor as the authors suggest that Hiaki prosody makes use of both stress and pitch accent. While it is unquestioned that the lexically contrastive pitch accent attaches to morae, it remains unclear whether the morphologically determined pitch accent does the same or attaches to syllables. Again, previous literature has argued in favor of either of the two (Dedrick & Casad 1999; Demers et al. 1999; Hagberg 2008). In order to resolve these questions, this study elicited data from a native Hiaki speaker and analyzed it acoustically in Praat (Boersma & Weenink, 2016). Measurements were taken for vowel length, pitch peak, and intensity peak. As it is not uncommon for a pitch accent to align relative to something within a word (e.g. Ladd et al. 2000; Bruce 2005;

[7] A detailed review in section 3.2 will outline why native English speakers were used as listeners and what the benefits of doing so are.

Ladd 2006), pitch peak alignment relative to the offset of the first vowel (V1) and as function of vowel length (V1) were also included.

2.1. Methods

2.1.1. Participants

The data for the acoustic analysis come from our Hiaki consultant Maria Levya. Maria was born and raised in Tucson, Arizona, primarily by her parental grandmother. Her grandmother was a very fluent Hiaki and Spanish speaker born in Yuma, Arizona. Since her childhood, Maria was exposed to Hiaki, Spanish, and English and is therefore trilingual. The recordings were elicited as parts of larger project on Hiaki at the University of Arizona, primarily documenting the language, compiling a learner's grammar, and understanding complex morphosyntactic issues. More information about those projects can be found at http://arizonahiaki.org/.

2.1.2. Stimuli

Words that only contained sonorous segments were chosen, ensuring a continuous pitch track across the entire word and avoiding microprosody caused by obstruents. A Hiaki dictionary (Molina et al., 1999) was searched for words that fulfill this requirement. Table 2 shows the non-inflected tokens included in this study and both the pronunciation and the meaning was confirmed by our consultants. *Waawa* 'roast in coals' is the only "all-sonorous" verb while all other words are nouns. *Moono* ‘doll’ and *miina* ‘mine’ are Spanish loanwords. Table 3 shows all inflected word forms that arose during elicitation and were included in this study.

Table 2: This table shows all the words used to elicit the stimuli.

Word	**Pronunciation**	**Translation**
naamu	/na:mu/	cloud
moono	/mo:no/	doll
naawa	/na:wa/	root
waawa	/wa:wa/	to roast in coals
miina	/mi:na/	mine

Table 3: This table outlines all affixed word forms which occurred in this study.

Word	**Translation**	**Word**	**Translation**
namu-ta	cloud-acc	mina-ta	mine-acc
mono-ta	doll-acc	mina-po	mine-at
moono-k	doll-prf	wawa-ne	roast.in.coals-fut
naawa-k	root-prf	nawa-tune	root-be-irr

2.1.3. Procedure

The words were elicited in weekly elicitation sessions with Maria and Santos Leyva but the data for this paper come exclusively from Maria. The Marantz Professional Solid State Recorder PMD620 was placed approximately 1m from the speaker’s mouth. Given the circumstances during elicitation, this was the best possible way to elicit data. It was constantly ensured that the microphone was facing our consultants with no interfering objects. The main goal was to control for as much as possible, be it linguistic interferences such as list intonation or extralinguistic interferences, e.g. slamming doors. If background noises such as airplanes or slamming doors interfered with the recording, the tokens were subsequently excluded. The target words were presented in the English translation to Maria, who would then ask Santos, most often in Spanish, to think of simple sentences in Hiaki that included the target words. Maria would spell out each word in the standard Hiaki orthography, the Latin alphabet, and once it was written on the board, she would read the sentences out a few times in a natural way. It was also attempted to make short dialogues out of the elicited sentences as it seemed that this would lead to a more natural, pause-free pronunciation. In other sessions, the target words were elicited in isolation by asking if Maria could tell us how to pronounce the Hiaki word for a given English word. Simply reading out items from a sheet of paper had failed in past attempts, so this was a better approach to elicit single words.

The resulting data included 47 tokens and were then analyzed in Praat. They come from pronunciations in isolation as well as from sentences. The sentences were either statements or questions. For each word, the maximum pitch and intensity were measured on the first and second vowel. The standard settings for the pitch tracker in Praat were not used; it produced many artifacts with this speaker's voice, perhaps because of her age and voice quality as well as her pitch range. The pitch range was set to 100-250Hz for most of the analysis. Most commonly, the voicing threshold was adjusted manually which made more of the pitch track visible. The octave-jump cost to was additionally set to a higher level as jumps were common. Both settings were usually changed by one to three tenths to achieve a stable pitch track. Altering those values by more than five tenths increased problems with the pitch track.

Sometimes, however, the pitch track would simply show the opposite of the acoustic impression, particularly in regions with creaky voice. In such cases, zooming in and manually measuring the frequency by spotting one cycle in the waveform usually confirmed the auditory impression so that the visuals were disregarded. This was done by identifying the erroneous part first. As it was unclear which part would include the pitch maximum, measurements at about 20%, 50%, and 80% were done manually. It was zoomed in until one could recognize one cycle of the waveform. The time for one cycle was measured from one valley in the waveform to the next valley. Praat automatically displayed the frequency for this cycle. The frequencies were compared and the highest was taken as peak maximum. The three measurements were additionally compared with one another to make sure that none of them was wrong, i.e. too dissimilar from the others. They were also compared to the beginning and end of the erroneous part; again, this was done to ensure that the hand measurements are correct and in line with the surrounding pitch contour.

Although it was necessary for most of the data to set the pitch range to 100-250 Hz to reduce pitch tracking errors, this setting caused pitch tracking errors in regions with unusually high pitch, such as the end of questions or continuation rises. In these cases, the pitch range was set to a higher maximum as needed for a given syllable. When segmenting the individual phonemes, a wideband spectrogram with an analysis window of 0.05ms was used to ensure good spectral resolution. For the pitch analyses, a narrow-band spectrogram with an analysis window of 0.025ms was used ensuring good frequency solution; this helped in getting a precise pitch calculation by Praat.

The data were then annotated, marking individual segments and the entire word. In most tokens, the border between individual segments was fairly obvious because most words consisted of nasal-vowel or vowel-nasal sequences. The boundary was then placed at the on- or offset of the F2. This point was usually very clear due to the low-energy anti-formant of the nasal. In cases of glides adjacent to vowels as in *waawa* ‘roast in coals’ or *naawa* ‘cloud’, the boundary was less clear. The boundary was then placed midpoint of the F2 transition.

Pitch peak on the first two vowels was measured and the same was done for intensity peak and vowel length. For both, the highest point within the syllable was located visually. All measurements were done manually. Vowel length was measured based on the segment boundary criteria described above. Measurements on the third syllable/fourth mora were not included since no literature to date would predict a pitch accent in such a position.

This study also includes measurements regarding the alignment of a high tone. Previous literature (e.g. Demers et al. 1999) has concluded that Hiaki realizes its pitch accent as a high tone on a particular mora or syllable. Extensive literature on other languages highlights that the alignment of a high tone's pitch peak to segmental landmarks is not uncommon and can convey pitch accent or stress distinctions; all this can furthermore interact with vowel length, e.g. Ladd 1999 and Ladd et al. 2000 for Greek, Japanese, and Dutch.

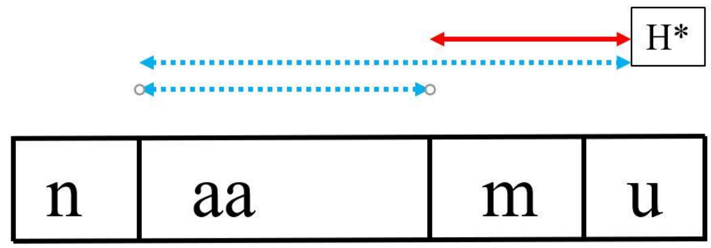

Figure 3: H* alignment measurements in terms of offset from V1 (the solid arrow) and as function of vowel length, i.e. the long dashed arrow divided by the short dashed arrow (= the length of V1).

The measurements presented in the following are outlined in figure 3. The first measurement, alignment offset V1, was measured as the length in milliseconds from the end of the first vowel to the pitch peak of this word. The measurement started at the end of the second formant for this vowel. This is highlighted in figure 3; the solid arrow illustrates this distance. The second measurement was the pitch peak alignment as a function of vowel length, accounting for possible differences between long and short vowels. First, distance from the onset of the first vowel to the pitch peak (i.e. the long dashed arrow) was measured and then divided by the length of the first vowel (i.e. the short dashed arrow). Note that the pitch peak for the alignment measurements was the highest pitch excursion on the entire word. The pitch on the first segment was not considered for this measurement as potential drops or rises due to the preceding word could have skewed the results. As the previous literature has furthermore highlighted, a pitch accent would be expected to occur on the either vowel but not on the first segment of a word.

2.2. Results

Although duration primairly tells us something about vowel length and not pitch accent, the vowel length measurements will be presented first as they might also be a cue for prominence in general. This will additionally verify that phonologically long vowels are also long phonetically and do shorten under affixation. Figure 4 illustrates the vowel length data. The y-axis denotes the length in milliseconds while the x-axis shows which vowels were measured. For instance, *naamU* means that this graph includes all measurements on the second vowel of words that contain a long first vowel and are not affixed; *nAAmu* represents the measurments for this long first vowel. *nAmuta* are all length measurements on the shortened first vowel under affixation while *namUta* represents all measurements for the second vowel in affixed cases. The boxplot demonstrates that long first vowels as in *naamu* 'cloud' or *waawa* 'roast in coals' are significantly longer than the second vowel in those words. This was confirmed by a one-factor two-levels within-subjects ANOVA, with the levels V1 for the first vowel, and V2 for the second vowel (this analyzes the data of the first two boxplots). The ANOVA was significant ($F(1,28) = 144.04$, *p* < .001). Another one-factor two-levels between-items ANOVA compared the first vowels in words with a long first vowel (***naa**mu*-cases)[8] and a short first vowel (***na**muta*-cases); this ANOVA was also significant (F(1,45) = 99.19, p < .001), comparing the second and fourth boxplot.

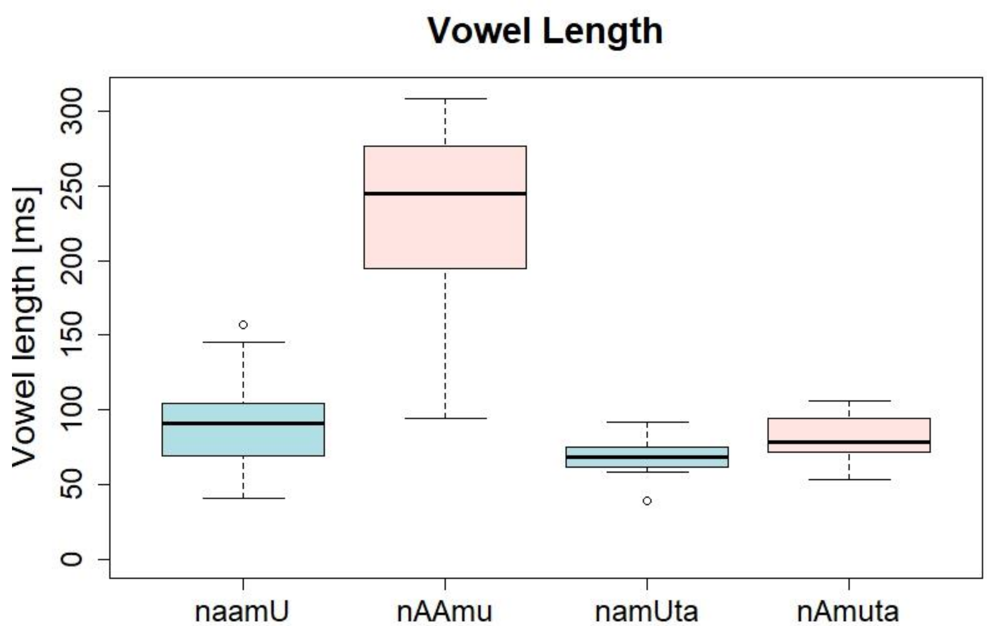

Figure 4: This boxplot compares the vowel lengths in different word forms. If a word has a long first vowel, those vowels (*nAAmu*) are significantly longer than the short ones in the same word (*naamU*) but also longer than the shortened first vowel under affixation (*nAmuta*). Capital letters indicate which vowel the boxplot refers to.

Turning now to measurements which directly reflect prominence, the pitch peak data will be illustrated in the following. The main question is whether there is a difference between the

[8] *Naamu*-cases refers to all unaffixed stimuli, i.e. *naamu* 'cloud', *naawa* 'root', *miina* 'mine' *waawa* 'roast in coals', and *moono* 'doll' (= all words in table 2). By contrast, *namuta*-cases refers to all affixed word forms of those five unaffixed ones, i.e. all words in table 3.

pitch on the first and second vowel. Figure 5 illustrates the data in form of a boxplot and a one-factor two-level within-subjects ANOVA returned a significant difference (F(1,28) = 19.59, $p < .005$) for instances containing a long first vowel (*naamu*-cases); this analyzed the two boxplots

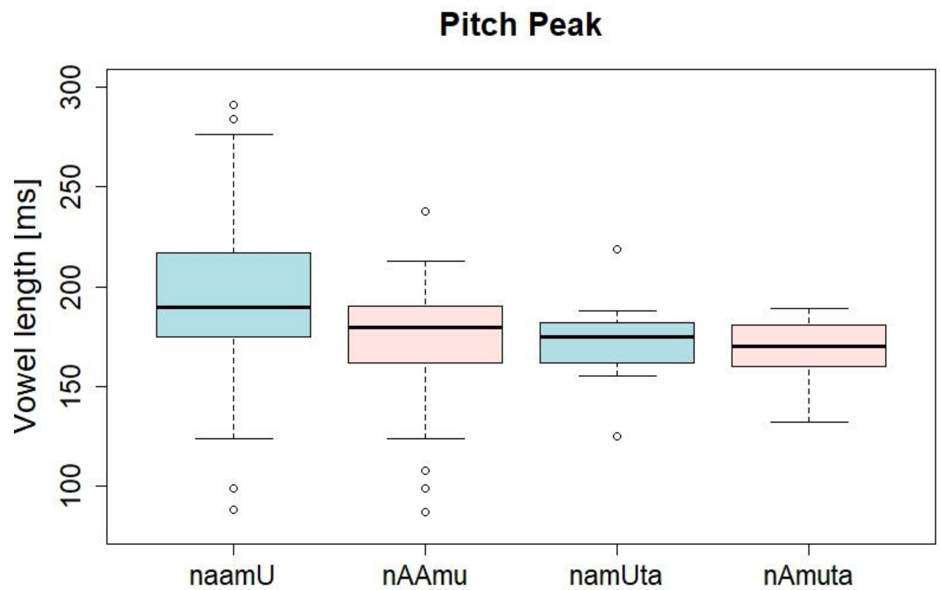

Figure 5: Pitch peak on the first and second vowel when the first vowel was long, e.g. in *naamu* ‘cloud’, or short, e.g. in *namuta* ‘cloud-acc’.

on the left asking whether the pitch peak is significantly different on long first vowels when compared to the second vowel in the same words, i.e. the peak on /a:/ versus the peak on /u/ in, for instance, *naamu* ‘cloud’. The spectrogram in figure 6 highlights this distribution. The word *miina* ‘mine’ is uttered in the sentence *uu miina pattawak* ‘the mine is closed’. There is a slight pause between *miina* and *pattawak* but listening to the entire sentence suggests that this is simply the transition from *miina* to *pattawak* in a rather slow speech rate, not a deliberate pause by the

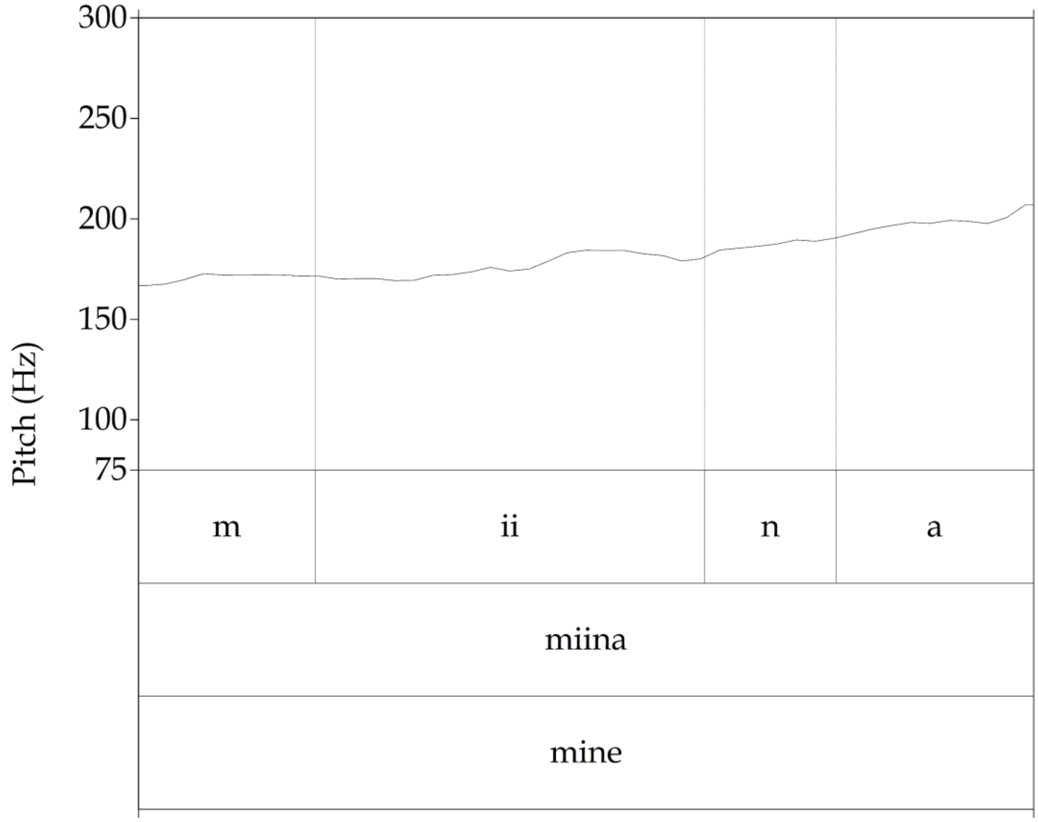

Figure 6: Pitch track of *miina* ‘mine’ showing the rise in pitch from the long first vowel /i:/ to the second short vowel /a/.

speaker causing a continuation rise. The spectrogram shows how the pitch peak is lower on the long first vowel compared to the second one. This is also what the first two boxplots in figure 5 show and the ANOVA returned as a significant difference. However, one might suggest treating this significant result with caution. The error bars, particularly for the second vowel (the leftmost boxplot) are very long and one can also observe several outliners for either vowel.

The data for the first and second vowel for tokens with a short first vowel, i.e. *nAmuta* and *namUta* instances, is displayed by the two boxplots to the right (figure 5); the data are almost identical. A one-factor two-level within-subjects ANOVA showed no effect ($F(1,17) < 1$). Looking at a sample spectrogram, one can see the boxplot distribution very well. Figure 7 illustrates the pitch across the word *monota* 'doll-acc' in a declarative sentence with no pauses; the sentence is

Figure 7: Pitch track of *monota* 'doll-acc' with not clear pitch peak on either of the first two vowels.

Esperensa aa asowarawa monota si tutu'ulik aa nu'uriak 'Esperansa bought her niece a really beautiful doll'. One can see that even though there is a slight pitch movement on the first segments, peak as well as the low on the first and second vowel are very similar. The contour itself seems to occur from the end of the first vowel /o/ over the nasal /n/. Both vowels do not have much of a contour nor a clear peak.

While intensity peak has not proven to be a cue for the pitch accent in Hiaki in previous studies, it was still included in order to determine whether it might be a characteristic of a pitch

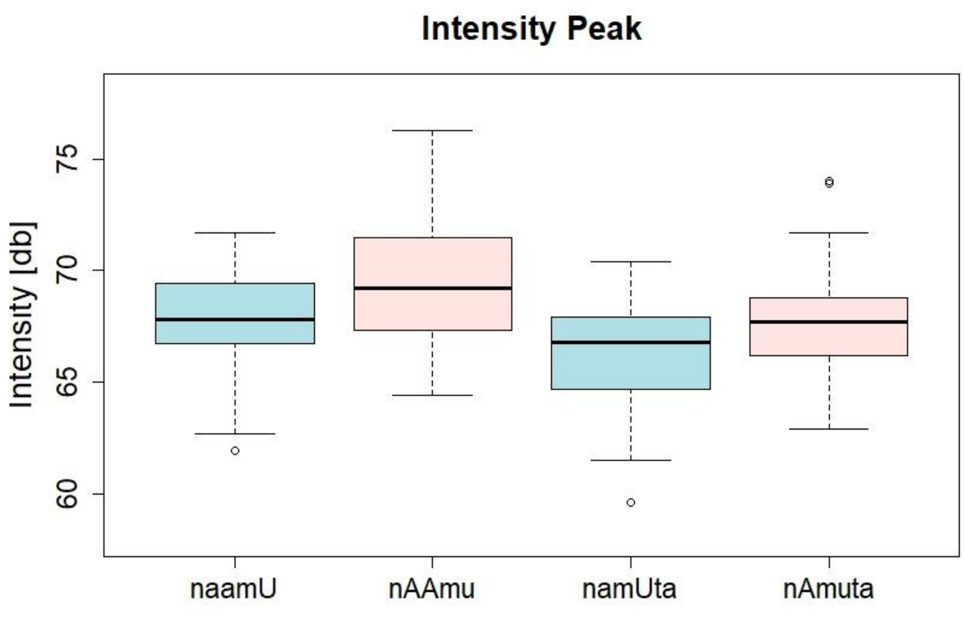

Figure 8: Intensity peak on V1 and V2 on unaffixed words (e.g. *naamu* 'cloud') and affixed words (e.g. *namuta* 'could-acc').

accent. The results for intensity peak mirror that of pitch peak. If there is a long vowel, the intensity peak between the first and second vowel returned a significant effect for vowel. It was tested with a one-factor two-level within-subjects ANOVA ($F(1,29) = 10.15$, $p < .005$). The intensity peaks did not return a significant difference when analyzing the difference between the first and second vowel in *namuta*-cases ($F(1,46) = 1.18$, $p > .2$).

Last is the possible alignment of a high tone, the pitch peak of the entire word. Figure 9 illustrates the alignment of the pitch peak after the offset of the first vowel, with the measurements given in milliseconds. The data between *naamu-* (long first vowel) and *namuta-* cases (short first vowel) look fairly similar with few outliers. A one-factor two-level within-subjects ANOVA did not show any effect ($F(1,45) < 1$). This demonstrates that the highest pitch tends to align at the same time after the offset of the first vowel. This alignment as a function of

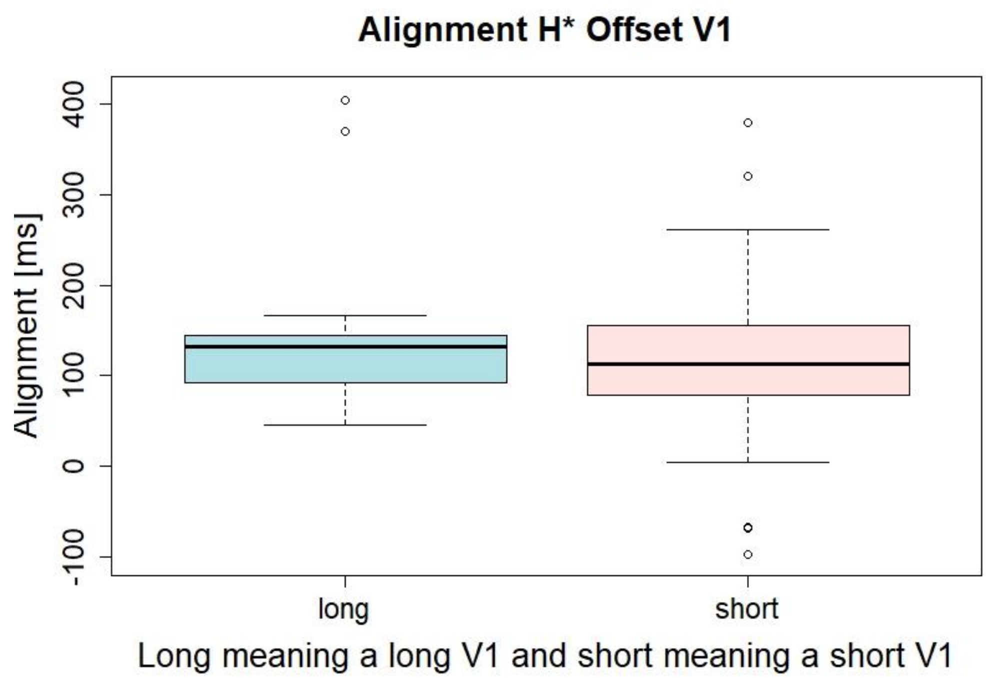

Figure 9: Alignment of the pitch peak after the offset of the first vowel.

vowel length showed a great dispersion. The data were therefore not considered any further as no conclusion is to be drawn in terms of alignment as function of vowel length.

2.3. Discussion

The only clear result emerging from the acoustic measurements is that phonologically long vowels are very long; if they shorten under affixation, they are significantly shorter. The pitch measurements do not provide any clear picture. There is no distinct peak on either of the two vowels in most tokens and the data exhibit a great distribution which makes it hard to understand it. The significant effects found though are that the pitch on first vowel in *naamu*-cases is significantly lower than on the second vowel and that the intensity is higher on the first vowel in *naamu*-cases. Based on those measurements however, it is difficult to draw any reliable conclusions about how the morphologically determined pitch accent is realized.

There may be three possibilities why no definite pitch peak for the pitch accent emerged: attrition, task-effects, or age. The reason why language attrition may play a role here is that Hiaki is only spoken by about 500 people in Arizona who, in the everyday life, will face considerable pressure from English and Spanish, the two main languages used in Arizona. Because of that, it could be the case that attrition might be one reason why no clear pitch accent emerged. Looking at the lexically determined pitch accent, this pitch accent does not seem to be a major feature of the language as it only distinguishes a small number of minimal pairs. If this low importance transfers to the morphologically determined pitch accent remains unclear. Yet, the fact that prosodic prominence may also be cued by other characteristic such as length or intensity, the fact that the lexically determined pitch accent only distinguished a small number of minimal pairs, and the pressure from English and Spanish could all work in concert to cause a morphologically determined pitch accent to attrite; this possibility will be discussed in the following.

De Leeuw (2009, p. 4-6) defines attrition as a non-pathological, not age-related loss of a native language within a single individual; this means that language loss is not related to a medical condition such as aphasia and does not refer to any that would naturally occur as an individual becomes older. Research on segmental attrition is common (Mayr et al., 2012; Rafat et al., 2017; de Leeuw et al., 2017) but research on attrition is scarce when it comes to intonation. Alvord (2006) investigated intonational patterns in several Cuban immigrant generations in Miami, Bullock (2009) analyzed focus in situ and the prosody of left dislocation, and de Leeuw (2011) looked into pre- and post-nuclear alignment in German bilinguals. All those studies showed that prosody can attrite and "[m]any linguists, especially those engaged in contact research, regard intonation to be particularly vulnerable to cross-linguistic influence" (Bullock, 2009, p. 168). It might therefore be not surprising that the consultants for this study may show effects of L1 prosody attrition, especially given that Hiaki has always been in close contact with Spanish and English. However, I would propose that this is not the case for our consultant. When listening to full sentences, the consultant produces a rich prosodic system, the target words included; this is, at least for naamu-cases, also true in isolation. Additionally, the consultant also exhibits a distinct lexical pitch accent on the minimal pairs mentioned earlier suggesting it might be task effect.

As just outlined, the reason why it seems hard to find clear pitch excursions on isolated words might be due to the task itself. When reading out or simply saying individual sentences, it is easy to observe a rich intonational system in the Hiaki recordings. This implies that the consultant's speech has (in sentential context) distinct intonational patterns which are far from non-existent or lost. This could suggest that it may be more of a task effect than L1 attrition. Generally, I would suggest that if a word would exhibit a pitch accent, that this would rather happen in isolation than in context. However, there has been no data or evidence that would suggest that it is not the opposite in Hiaki; it might very well be the case that Hiaki does not realize its morphologically determined pitch accent in isolation.

Last, age might also play a role. One could surmise that the pitch range may shrink in older adults so that the data presented here will not show distinct peaks. However, it is unlikely that those results stem from age alone. Dupuis and Pichora-Fuller (2010) demonstrate that older adults can comprehend and produce affective prosody in the same way as young adults do. Stathopoulos et al. (2011) illustrate that the F0 variability for older adults is greater than for younger ones. Considering F0 range, McGlone and Hollient (1963) could not detect an age-based differences. Still, Ptacek et al. (1966) found a smaller F0 ranger in older adults. The researchers used a sustained vowel and a monosyllabic word production task. Barnes (2013) reports that those conflicting results could be due to the task alone. Keeping this in mind for the data here, saying words such as *naamu* 'cloud' or *moono* 'doll' in isolation may then illustrate what Ptacek et al. (1966) discovered: a reduced pitch range as a task-effect. This could then provide some insights why it was particularly tokens in isolation that exhibited a near flat pitch track. Because of the language's endangered status, it is not possible to record additional younger fluent speakers.

Considering the intensity measurements, one cannot reach strong conclusions. First, the intensity between vowels in cases with a long first vowel is very often identical and the same goes for the overall distribution across all tokens. This suggests that wherever a pitch accent is realized on affixed words such as *namuta* 'cloud-acc' is not influencing intensity. It might also be the case that whatever causes the flat pitch to occur may also apply to intensity showing no differences between the vowels. One must also remain skeptical of the significant outcome in *naamu*-cases as the data still show a great dispersion with long error bars. In addition, intensity varies quite easily. In a short review, Bogomolets (2014, p. 59) highlights that intensity tends to be a rather unreliable cue for stress or pitch accent. Intensity is furthermore quite susceptible to noise, the consultant's head movements, interfering objects, the distance to the microphone, and other environmental factors. Bearing this in mind and looking at the raw values, which display little variation, I would suggest that it remains difficult to draw reliable conclusions from this data. We get a hint that intensity on the long first vowel in *naamu*-cases is significantly greater than on the second vowel but considering the elicitation method and the dispersion in the data, one must remain cautious to base interpretations solely on intensity.

Next is pitch peak. Fabián and Duarte (2006) and Hagberg (2008) claim that F0 is a reliable cue for the Hiaki pitch accent. The peak measurements presented here do not agree with this research and no conclusion can be drawn. However, a closer look at the data reveal an interesting finding. When analyzing the data, many tokens exhibited a slight fall starting after the initial segment followed by a continuous rise over the rest of the word (see figure 10). Sometimes, this rise would manifest in a curve but most often, it remained a constant rise. While this was the impression on a rather uncontrolled data set which was not used for this paper, it still looked like that the tokens presented here had a similar pattern. Figure 10 is such an example. The pitch contour illustrates how the pitch tracker dips into the first vowel before it rises again over the course of the following two segments. This pattern additionally occurs in tokens with a shortened vowel and even extends into the following segments after the second vowel. Therefore, the pitch on the onset of V1 was re-measured and compared to the pitch peak of the entire word. The same goes for the lowest and highest pitch on either vowel and the contour was noted. This gave rise to a pattern presented in table 4. Note that the total numbers are not the same because four V2-contours were bell-shaped and excluded. If the first vowel is long, the tokens predominantly show a rise on either vowel whereas if the fist vowel is short, we

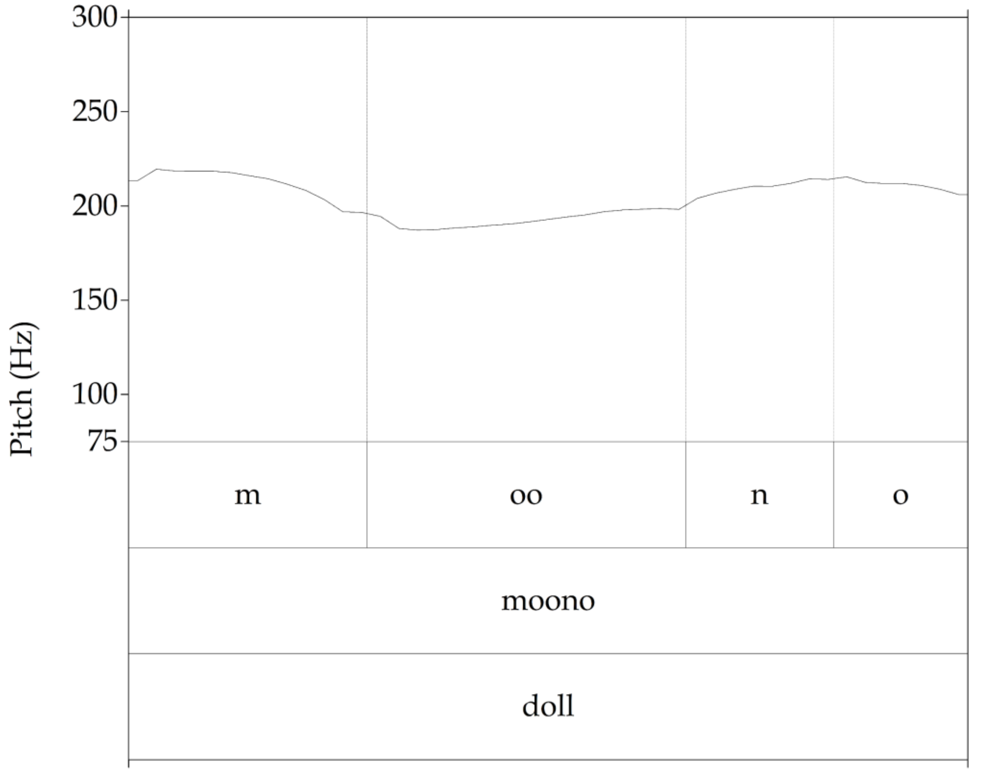

Figure 10: Pitch track of *moono* ‘doll’, first dipping into the long vowel [o:] before slightly rising towards the end of the word.

Table 4: Overview of the contours on the first and second vowel depending on the

Vowel Length	**V1 Contour**			**V2 Contour**		
	total	**rise**	**fall**	**total**	**rise**	**fall**
Long (*naamu*)	29	20	9	28	26	2
Short (*namuta*)	18	4	14	15	13	2

see the opposite pattern for vowel length of the first one V1 and V2. A shortened V1 usually exhibits a fall while the second vowel then shows a rise. This reflects the noticed pattern when analyzing the data and offers an interesting pattern. One might assume that this could be a result of the morphological process as the pattern arose from tokens in isolation, statements, and questions all together. However, we can only observe 57 tokens so that the next step should be analyzing further tokens to find out whether this pattern will hold. Another limitation comes from the raw values. About one third of all contours in table 4 shows an intra-syllable pitch difference of 10Hz or less. This could be, as outlined earlier, a result of age and therefore still be important but it might also suggest that those rises are too small to be perceptually salient. A possible answer to this is experiment 2.

3. Experiment 2

The acoustic measurements in experiment 1 have not been provided a clear picture of the morphologically determined pitch accent so that experiment 2 will now address this question from a perceptual perspective. The main question is: which syllables of the target words from experiment 1 do listeners perceive more prominent? Considering the measurements earlier, I would predict that syllables with a long vowel (e.g. ***naa****.mu* 'cloud') are perceived more prominent since those long vowels are often up to three times longer than the short ones; this is likely to happen in isolation as well as in statements. In affixed word forms such as *namuta* 'cloud-acc', it remains unclear which syllables listeners will mark as prominent. I expect to see higher prominence ratings on target words in isolation than in statements because they lack a long vowel which I assume will stand out of the speech stream easily. In addition, sentence intonation patterns come into play that may reduce the prominence of a potential pitch accent. Whether listeners mark the first or second syllables in tokens such a *namuta* 'cloud-acc' as more prominent remains to be seen but will then answer one question, namely does a potential pitch accent shift under affixation or not. If the pitch accent shifts, then I would expect higher prominence ratings on the second syllable. If the opposite is the case, ratings should be higher on the first vowel.

Experiment 2 will now test non-native listeners with the RPT method (e.g. Mo et al. 2008; Cole et al. 2010; Cole et al. 2016) on Hiaki words in isolation and sentences. The sections below will outline the RPT method in detail and also provide a thorough explanation why native English speakers were chosen over native Hiaki speakers.

3.1 Participants

All participants came from the University of Arizona community and were native speakers of English. Most of them were either undergraduate students who received course credit or graduate students who participated voluntarily; none of the graduate students were in linguistics and none of the subjects reported a hearing problem or had any prior experience with Hiaki; some subjects spoke Spanish, French, or Arabic as a second language. The study included 36 participants. Cole et al. (2017, p. 322) report that studies with native speakers require about 10-12 listeners to produce reliable results when examining prosody; this depends on the situation and the materials tested so that in some cases, as few as seven speakers suffice.

The reason why this study included native English speakers listening for prominence and not native Hiaki speakers stems from several circumstances. One reason is that access to native Hiaki speakers other than our consultants was rather limited and not easy to achieve; in addition, dialectal differences between speakers from Arizona and Sonora exist. It is unknown how those differences would affect pitch accent placement and the prosodic perception thereof because it remains unclear whether there are pitch accent differences between the two dialects.

Another reason for choosing non-native speakers over native speakers is that previous research has demonstrated that L2 speakers can very well mark prominences in a foreign language. Printer et al. (2014) compared English native speakers with Japanese learners of English. While ratings from English and Japanese listeners were not identical, they were still highly similar. Printer et al. (2014) state that for prominence scores, "there is also a strong correlation between native and Japanese EFL speakers [...] ($r = 0.749$, $p < 0.01$)" (p. 546). Figure 11 depicts this strong correlation. Printer et al. (2014) also note that the overall results between native English and non-native Japanese speakers show similar patterns in form of a correlation between prominence scores and acoustic cues measured; the authors isolated vowel length as the strongest predictor for both groups. Unexpected is also their result that pitch-related acoustic cues did not dominate for the Japanese listeners, a pitch accent language. To conclude, their study demonstrates that non-native speakers can perform prominence marking at a similar level as native speakers.

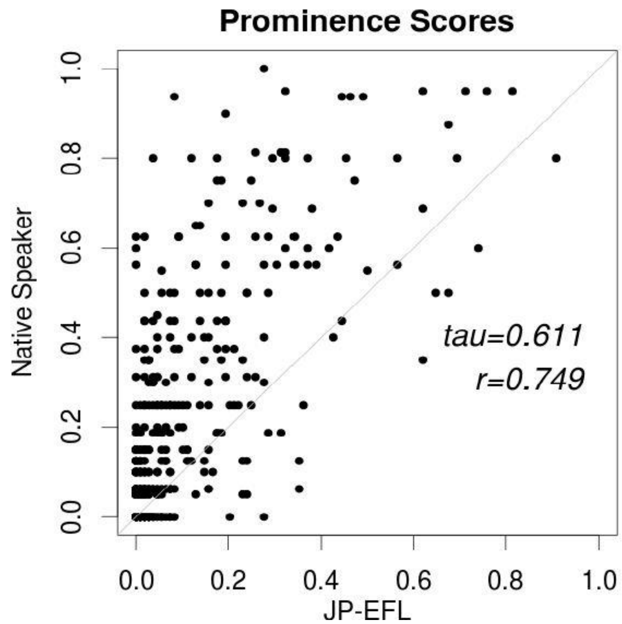

Figure 11: Correlation of prominence scores between native and non-native speakers (Printer et al., 2014, p. 546).

Furthermore, non-native speakers do not have any biases, e.g. from the lexicon, that could arise being a native Hiaki speaker. Cole et al. (2010) analyzed the effect of acoustic as well as discourse cues on listeners marking prominence in the Buckeye Corpus (Pitt et al., 2007). Their study highlighted that it is not only acoustic cues that lead listeners to perceive prominence; word frequency and repetition play a role as well, not as strong as duration and overall intensity but they still contribute. The researchers showed that highly frequent words, e.g. function words, are often highly reduced and consequently not very prominent. The same goes for words which are repeated throughout a conversation. The more often a word is uttered, the more reduced speakers will produce it later on. In an analysis that excluded function and highly reduced words, the authors still found a highly significant effect of word frequency. The study here avoids all those potential effects by choosing native English over native Hiaki speakers; those non-native speakers can then solely rely on the mere acoustic cues present in the signal while contextual or frequency factors are excluded; this way, lexical biases are also excluded.

A further deviation from previous studies is that the sentences were split into syllables visually on the response screen because marking the prominent words in a sentence would have not addressed the research question. In other words, non-native speakers were required to mark prominent syllables over the course of entire sentences as the question is which morae or syllables are prominent, rather than which words. The study included 36 listeners. Admittedly, there is no assurance that 36 subjects will produce reliable results and there is no literature that would help this case. Yet, keeping Cole et al.'s 2017 study in mind, even listeners who were presented with an unfamiliar variety, 36 participants were almost always in the range of approximately 0.04 standard deviations and the general trend seems to be a falling exponential curve so that 36 listeners for this study should be sufficient to achieve a reliable result.

3.2 Stimuli

The stimuli are the same as in experiment one. They were either single words uttered in isolation or short sentences. When the elicitation sessions provided relatively long sentences such as

Vasiilia hunuka huya nawata bwa'aka ko'okoe wechek 'Vasiilia ate that plant root; she fell sick' for the example *namuta* 'cloud-acc', those sentences were cut into shorter stimuli at intonational boundaries. Pilot tests showed that it was extremely difficult for listeners who do not speak Hiaki to find prominent syllables in utterances that long. They were therefore cut short and only parts of the entire sentences were presented. This should have little to no influence on the research question because within each intonational phrase, the target would still be exhibited to sufficient sentential intonation. Subjects listened to 55 stimuli total; eight stimuli were shorter portions of sentences.

3.3. Procedure

The listeners were seated in a sound-attenuated booth in the SPAM-Lab at the University of Arizona. The stimuli were presented over ATH-M40fs Audio Technical headphones at a sound level comfortable for each participant. After signing the consent form, each subject received detailed instructions on how to proceed. Those instructions were again presented on the screen. It was highlighted several times that there is no right or wrong answer to this experiment and subjects should simply listen and click on whichever syllable they heard as more prominent. The experiment used LMEDS, the Language Markup and Experimental Design Software (Mahrt, 2016a), which was specially developed to collect data for RPT (Cole et al., 2016; Cole et al., 2017). The following description is mainly based on Cole et al. (2017, p. 301-306) unless otherwise noted.

In most intonation research, trained listeners or experienced linguists annotate a language’s prosodic system, often with the Tone and Break Indices system ToBI. Those “lab-annotations” are reliable, often exhibit little inter-annotator variation, and difficult cases can be solved by talking to each other and finding an agreement. With the help of ToBI, specific acoustic and perceptual cues can be identified, but this leaves out contextual factors, e.g. syntactic factors, which can also have an impact (Cole et al., 2017, p. 302). In addition, ToBI works best for languages which are well understood since ToBI transcribes what is distinctive in the intonational system of a given language, not simply where peaks and valleys are. This leaves out those languages whose prosodic system is undocumented or remains a mystery. Such expert transcription is also costly and time-consuming.

A faster and cheaper way is RPT. The Rapid Prosodic Transcription method does not require any trained personnel. Naïve listeners who have no annotation experience listen in real-time to the language of interest. They are then asked to mark prominent words and often intonational boundaries as well. Subjects receive no guidelines or explicit criteria and instructions are kept minimal. The length and type of instruction can vary, depending on the task and research question (Cole et al., 2016, p. 8). Listeners work independently so that problematic cases cannot be resolved as with trained annotators in the laboratory and naïve listeners do not receive any feedback; they also have no supervision during the task. It is therefore important to tell participants that there is no correct answer and that differences between individual annotations are in fact a great source of information (Cole et al., 2016, p. 8). RPT is a real-time task so that annotators have little control over the stimuli which are always played in their entirety, rather than a phonetician's common method of selecting small portions to listen to separately. Most often, subjects are allowed to play the stimuli a second time to check their choices against the audio and make adjustments when necessary. Therefore, the basis for their annotation is their initial auditory impression. The transcript accompanying the auditory stimuli is presented without any capitalization as this might shift subjects' focus to particular words; the transcript is also free of punctuation marks and presented in normal orthography.

Obviously, this multi-annotator approach faces the problem of consistency and accuracy, but the total number of listeners marking any given syllable as prominent produces a continuous scale measurement of how prominent a given syllable is, rather than a categorical measurement of prominent vs. non-prominent. Compared to professional trained lab-annotators, naïve listeners start off with a disadvantage in form of a massive lack of knowledge and/or training. And yet, this approach has proven to be quite reliable (Mo et al. 2008; Cole et al. 2010; Cole et al. 2016; Cole et al., 2017). The strength of RPT lies in the fact that it uses not only one or two annotators but several. While each individual listener may attend to specific acoustic or contextual cues, a large group of untrained participants will indeed produce a reliable result, even in a foreign language. The cohort size, the number of listeners required to obtain “annotations [which] will be minimally sensitive to specific annotators performing the task” (Cole et al., 2017, p. 308), varies from task to task and from language to language but can be as low as seven participants. Nevertheless, Cole et al. (2017, p. 320) have shown that less than five subjects produce unusable results. RPT is particularly beneficial when exploring the prosodic system of a language that has not been well-documented, for instance Hiaki.

The experiment here differs slightly from the common RPT methods as just outlined. First, non-native speakers were tested on a language they have never heard because of the constraints of an endangered language and the potential native language interference such as lexical biases. Second, based on the research question, the data were split into syllables visually and subjects were asked to mark prominent syllables not words. The listeners were allowed to listen to the audio stimuli three times because pilot runs have shown that it was not an easy task to mark syllables in a language one has never heard when confronted with short stretches of connected speech. In addition, the accompanying transcript looked unusual as syllable boundaries were marked with a “.”. This is illustrated in figure 12, which shows the response screen for the stimulus *Merehilda namuta vichak* ‘Merehilda saw the cloud’. Participants received an English example as a practice item to familiarize them with how sentences would be displayed on the screen[9]. While it has been demonstrated that the instructions given can influence the result, this experiment made use of "neutral" instructions so that it would not bias participants to listen for something particular (Mahrt 2016b). The instructions were: "In normal speech, speakers pronounce some word or words in a sentence with more prominence than others. The prominent words are in a sense highlighted for the listener, and stand out from other non-prominent words. Your task is to mark words [*syllables* in this study] that you hear as prominent in this way" (Roy et al., 2017, p. 6). The experiment concluded with a short questionnaire asking subjects about their language background.

Figure 12: A sample screenshot of the experiment showing the sentence *Merehilda namuta vichak* 'Merehilda saw a cloud'.

The output is coded in a binary way, zero representing no prominence and one standing for a syllable marked as prominent by a listener. The total scores thus reflect the likelihood of a

[9] The English example presented was *in.ter.na.tion.al.is.an.ad.jec.tive* for *international is an adjective*.

syllable being heard as more prominent across the group of listeners, or the proportion of listeners hearing it as prominent.

3.4 Results

The data were compiled from 36 native English speakers who listened to each of the 55 stimuli once; it was analyzed with a 2-factor 2-levels within-subjects ANOVA. One factor was vowel length, i.e. short (*namuta*-cases) standing for all tokens that occurred in affixed versions with a short first vowel and long (*naamu*-cases) meaning all tokens that had a long first vowel and therefore did not have any affixes attached; the other factor was context, i.e. whether tokens were uttered in isolation or within a sentence. The dependent variable was the prominence score of all subjects on either the first or second vowel. Since the coding for each subject was binary, i.e. "0" for not prominent or "1" for prominent, it was averaged over items. Two ANOVAs were run separately; one for the prominence scores on the first vowel with the factors length and context and one for the prominence scores on the second vowel with the same factors.

The ANOVA for the ratings on the first vowel returned significant main effects for both length ($F(1,35) = 201.06$, $p < .001$) and context ($F(1,35) = 31.05$, $p < .001$). The interaction between the two factors was also significant ($F(1,35) = 13.16$, $p < .001$). The ANOVA for the ratings on the second vowel did not return any significant main effects (length: $F(1,35) = 1.93$,

$p > .1$; context: $F(1,35) = 3.14$, $p > .05$)). The interaction between length and context was significant ($F(1,35) = 52.22$, $p < .001$).

Consequently, the data for both V1 and V2 were split over context because the main interest is whether the prominence rating differs between long and short vowels, i.e. between unaffixed and affixed versions. The split data sets contain one set with all long and short tokens in statements and one set with all long and short tokens uttered in isolation. A 1-factor 2-levels within-subjects ANOVA with the factor length and the levels long and short was run on both data sets to test for simple effects. Considering V1 first, the simple effects for both isolation ($F(1,35) = 149.87$, $p < .001$) and statements $F(1,35) = 184.37$, $p < .001$) were significant. Looking at V2, again, both simple effects for isolation ($F(1,35) = 21.32$, $p < .001$) and context ($F(1,35) = 16.76$, $p < .001$) were significant.

The data is displayed in figure 13a and figure 13b. The y-axis shows the prominence rating where zero stands for not prominent and one for prominent. The higher this score, the more listeners rated those tokens as prominent. In other words, if all subjects marked all tokens in a specific condition as prominent, the score would be one. If no subject marked any token in this specific condition as prominent, the score would be zero. The findings which will now be presented are also summarized in table 5 below.

Table 5: Overview of the simple effects split into V1 and V2

V1	**naa**mu	more prominent	isolation/statement
	namuta	less prominent	isolation/statement
V2	naa**mu**	less prominent	isolation
	na**mu**ta	more prominent	isolation
	naa**mu**	more prominent	statement
	na**mu**ta	less prominent	statement

Figure 13a illustrates that in isolation, the long first vowel in, for instance, *<u>naa</u>mu* ‘cloud’ (boxplot *long.isolation*) is perceiveid significantly more prominent as the short first vowel in, for

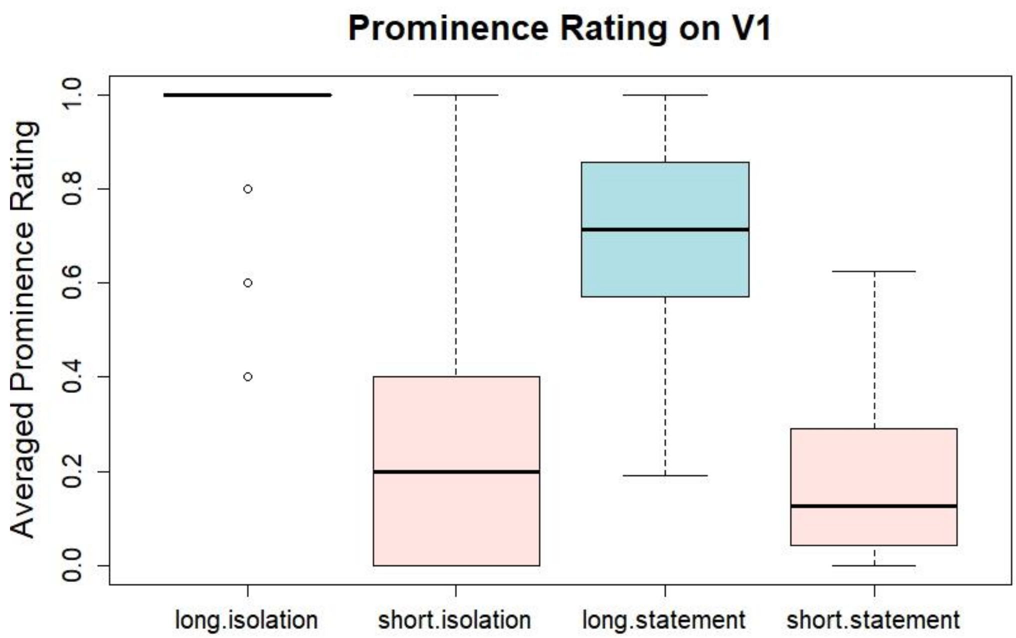

Figure 13a: This figure shows the prominence ratings on the first vowel for cases with a long first vowel (*naamu*-cases) and short first vowel (*namuta*-cases) in both contexts, isolation and statement. For example, *long.isolation* means that this boxplot is comprised of all **naamu**-cases that were uttered in isolation. The y-axis indicates the prominence score. The higher this core, the more prominent the tokens were rated.

example, *na*mu*ta* 'cloud-acc' (boxplot *short.isolation*). The same is true in statements. Again, the long first vowel in *naamu*-cases (boxplot *long.statement*) is perceived significantly more prominent as the first short vowel in *namuta*-cases (boxplot *short.statement*).

Looking at figure 13b for the prominence ratings on V2, we can see a difference between words in isolation and words in statements. In isolation, V2 is perceived significantly more prominent for *namuta*-cases than for *naamu*-cases. In other words, the /u/ in *namuta* 'cloud-acc' is more prominent than the /u/ in *naamu* 'cloud'. Figure 13b highlights this result with the first

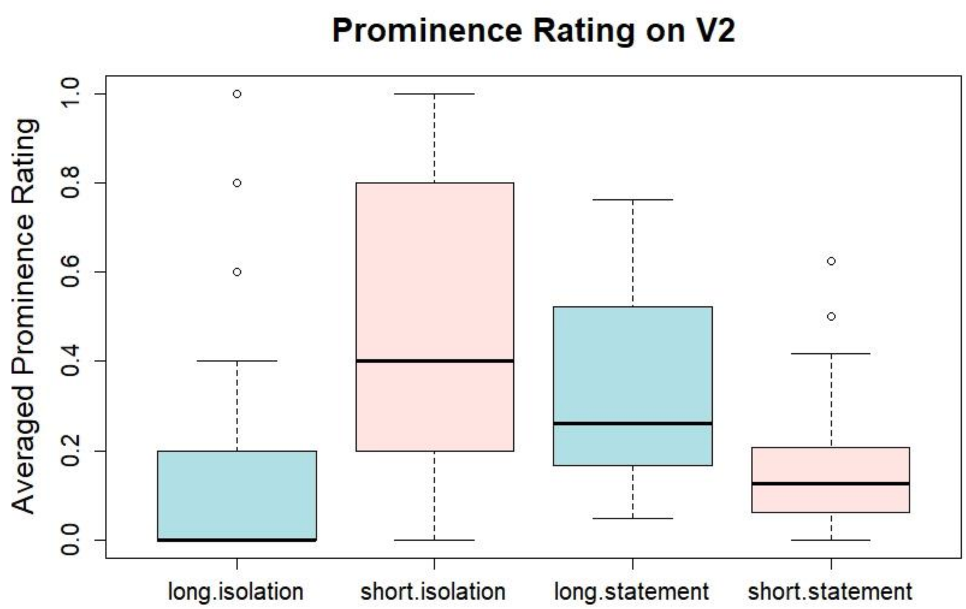

Figure 13b: This figure shows the prominence ratings on the second vowel for cases with a long first vowel (*naamu*-cases) and short first vowel (*namuta*-cases) in both contexts, isolation and statement. The y-axis indicates the prominence score. The higher this core, the more prominent the tokens were rated.

two boxplots from the left, i.e. *long.isolation* and *short.isolation*. However, the opposite is the case for V2 in statements. In *naamu*-cases, V2 is perceived significantly more prominent in statements than V2 in *namuta*-cases.

3.5. Discussion

The results from experiment 2 show that both in isolation and context, the long vowels in *naamu*-cases received high prominence ratings, almost at ceiling for tokens in isolation (see Figure 13a, boxplot *long.isolation*). The opposite is the case for the second short vowel in those *naamu*-cases as these instances received very low ratings. More than 50% of the listeners rated them as not prominent at all (see Figure 13b, boxplot *long.isolation*). This outcome is not surprising given the extreme length of those long vowels, as experiment 1 has illustrated. When heard in isolation, subjects listened to only those words, i.e. two syllables; one of those syllables has then a vowel which is often three times longer than the short second one. In addition, the intensity measures from experiment 1 also showed that the long first vowels in *naamu*-cases are uttered with significantly higher intensity than V2 in those *naamu*-cases. Listeners therefore behaved as expected and marked the long vowels in isolation as more prominent than the short ones. This pattern upholds in statements to a lesser degree (compare boxplots *long.statement* for V1 in Figure 13a with *long.statemen*t for V2 in Figure 13b). Again, this is not surprising because listeners were faced with several words at the same time, connected speech features, sentence intonation, and all that in a language they did not speak. It could therefore be considered normal to see slightly lower prominence ratings for *naamu*-cases in statements but also in general. Still, the pattern is the same: the long vowel receives significantly higher ratings than the short one. This might be due to the reasons just mentioned. In addition, length is also a cue for stress in English so that the non-native listeners in this experiment were presented with a cue for stress in their native language (cf. Gussehoven, 2004).

The situation is different for the V2 ratings. As expected, the overall ratings are lower for *namuta*-cases in statements than in isolation. The reasons were just outlined. Simply said, a lot more things are going on in statements that could have grabbed the listeners' attention than in isolation. This is accompanied by the absence of longs vowels. Ratings on *namuta*-cases in isolation are therefore higher in general. The pattern for those *namuta*-cases, however, is slightly different and for the two contexts statement and isolation. In isolation, V2 received slightly higher ratings than V1, i.e. listeners were more likely to mark prominence as *namúta* 'could-acc' than *námuta* 'cloud-acc'. In statements, the median in *namuta*-cases is the same for both vowels. The distribution of data shows that V1 received a few more higher ratings than V2 but there is not much difference overall. The fact that there is little difference for all *namuta*-cases in statements is likely due to the reasons outlined above.

Overall, listeners behaved mostly as predicted and prominence ratings were generally higher in isolation than in statements. They rated instances with a long first vowel as significantly longer regardless of context. In cases of a short first vowel, the ratings vary and depend on the context. In isolation, the second vowel (e.g. *namuta* 'cloud-acc') received higher prominence ratings that the second one in *naamu*-cases but also higher ratings than the first short vowel in the same token (e.g. *namuta* 'coud-acc'). Considering statements, the second vowel in *namuta*-cases received lower scores than the second one in *naamu*-cases but those ratings were very similar to the first vowel in the same token.

4. General Discussion

4.1. Summary of experiment 1 and experiment 2

The present work investigated the morphologically determined pitch accent in Hiaki. Two initial research questions guided this study: what are the acoustic correlates of this pitch accent and does this pitch accent attach to syllables or morae. Experiment 1 addressed the first question and included vowel length, pitch peak on the first and second vowel in word forms such as *naamu* 'cloud' and *namuta* 'cloud-acc', intensity peak, as well as pitch peak alignment as raw value from the offset of the first vowel and as function of vowel length. Those measurements were included in order to cover a wide range of possible prominence cues. As those acoustic measurements in experiment 1 remain slightly inconclusive, a second experiment was included attempting to address the question about acoustic correlates from a perceptual side. The idea behind this experiment was that if listeners mark certain syllables as prominent, certain acoustics would have to cue this prominence so that one could investigate whether the perception experiment correlates with some acoustic measurements. Thirty-six native English speakers listened to 55 Hiaki words in isolation and sentences and marked prominent syllables with the RPT method. The results from both experiments are summarized in table 5 as a short overview.

Table 5: Summary of the results from both experiments

Experiment 1	Vowel length	Phonologically long vowels are also phonetically long (up to three times as long as the short ones)
	Pitch peak	Significant difference between pitch peak on *naamu*-cases (lower pitch on V1)
	Intensity peak	Significant difference for intensity in *naamu*-cases (greater intensity on V1)
	H* alignment (raw)	H* aligns similarly when looking at the raw milliseconds from the offset of V1
	H* alignment (ratio)	H* does not align as function of vowel length
Experiment 2	Prominence Ratings	(1) Long vowels in *naamu*-cases received the highest ratings regardless of context (2) In isolation, V2 in *namuta*-cases received higher ratings than V1 in such tokens or V2 in *naamu*-cases (3) In statements, V2 in *naamu*-cases is more prominent than V2 in *namuta*-cases with no difference between V1 and V2 in *namuta*-cases

4.2. The acoustic and perceptual correlates

Combining the results from both experiments, it remains difficult to highlight the precise acoustic correlates of the morphologically-determined pitch accent. Considering the *naamu*-cases first, the acoustic measurements showed a significant intensity difference implying greater intensity for the first vowel. The measurements also highlighted a significant difference in pitch, with a lower pitch on the first vowel and a higher pitch on the second vowel. Last, the length measurements confirmed that those phonologically long vowels are significantly longer than short ones and English-speaking listeners strikingly marked those long vowels with a lower pitch and greater intensity as significantly more prominent than V2 in all *naamu*-cases. Those tokens then fit into Fabián and Duarte's (2006, p.7ff) analysis. The authors suggest that *naamu* 'cloud' has a single aligned iambic foot [W S] in that the pitch will peak on the second syllable, just as it did in experiment 1; the [W S] foot will therefore align with [L H] tones. However, they also find an intensity peak on V2 which stands in contrast to experiment 1, where the peak occurred on V1. I would suggest to treat intensity peaks with caution. As Bogomolets (2014) writes, intensity measurements can vary very easily so that thinking about the elicitation of the tokens for experiment 1, that could have been the case since the distance between the speaker and microphone was about 1m. Yet, since intensity was compared within items, the chance that our consultant moved their head while saying a single word seems unlikely. Considering Fabián and Duarte's (2006) intensity claims, the authors occasionally present differences of less than 1db as a crucial difference which is, given how easily intensity can be influenced during elicitation, not a lot. The two also state that length does not play any role for stress placement as words with shorter vowels than *naamu* 'cloud', e.g. *baakot* 'snake' (on average, the researchers note that those long vowels are 70ms shorter than the ones in *naamu*-cases), or words with a short first vowel, e.g. *yeka* 'nose', also follow the pattern of a single-aligned iambic foot (Fabián & Duarte, 2006, p. 9). While the naamu-cases from experiment 1 do not entirely fit into what Fabián and Duarte (2006) found, it still exhibits similarities. We will now turn to the acoustic measurements for all *namuta*-cases.

The acoustic measurements for *namuta*-cases provide little insight because none of them exhibits a difference between the first and second vowel. The vowels have roughly the same length, the pitch peak is fairly identical, and so are the intensity peaks. Looking at the results of the perception study, the ratings are generally lower than for *naamu*-cases. In isolation, V2 is rated more prominent as V1 with only minor differences between the two vowels in statements. However, a more detailed look is desirable here. Figure 14a depicts the prominence ratings on the first, second, <u>and</u> third vowel of all *namuta*-cases uttered in isolation. What one can see is that those ratings increase throughout the word, resulting in V3 being more prominent than V1

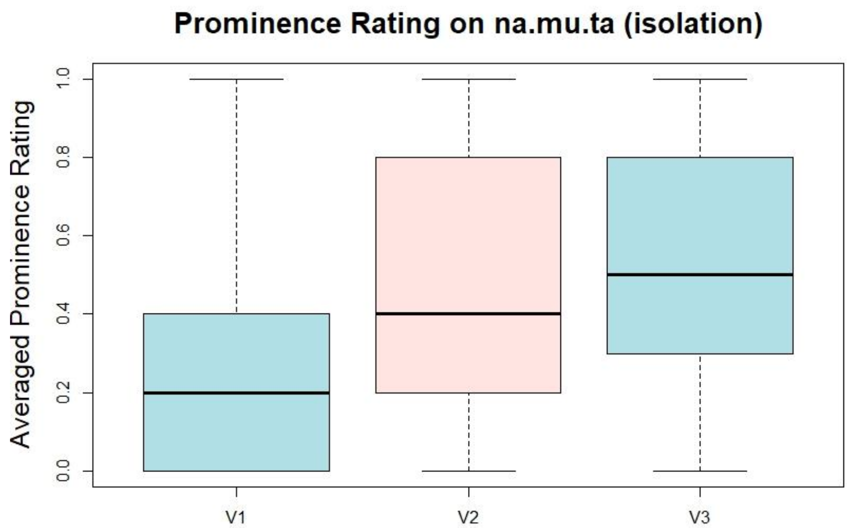

Figure 14a: Prominence ratings for *namuta*-cases on V1, V2, and V3 in isolation

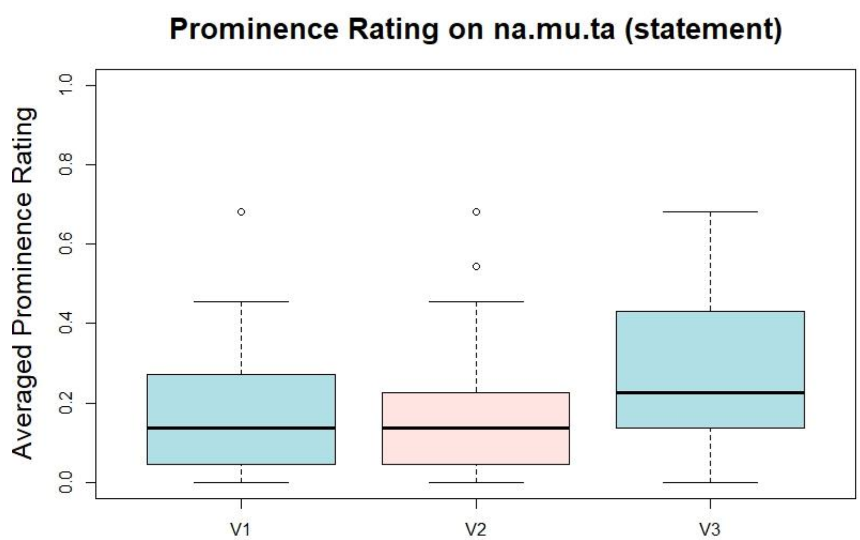

Figure 14b: Prominence ratings for *namuta*-cases on V1, V2, and V3 in statements

and similarly prominent to V2. The graph also illustrates that the ratings assumed all possible values from 0 to 1 and that native English-speaking listeners marked prominences rather towards the end of the word. Figure 14b shows the same ratings for the same tokens but in statements. Apart from the fact that the ratings are comparably low in general, the pattern also looks slightly different. While the ratings on V1 remain low, the ratings for V2 are now similar to that of V1 whereas V3 shows higher ratings (it is not surprising that the ratings are generally lower than in isolation since more things are going on that could distract non-native listeners from marking prominent syllables on the target words; on top of that, listeners listening to statements were not forced to mark any syllable on the target words as prominent, a decisive difference when listening to targets in isolation). Judging from those ratings, one might conclude that prominence is located either on V2 or V3. The acoustic measurements on V1 and V2 exhibit almost no differences and measurements for V3 were not included as no previous study has found prominence on the third vowel. The reason for the outcome in isolation in figure 14a might then be due to the fact that listeners were forced to pick at least one syllable as more prominent. LMEDS would participants only allow to proceed if one or more syllable was marked as prominent. Given that there might not have been any significant difference in a measure related to prosody, subjects could have randomly picked a syllable. This argument can be supported by the mere size of the boxplot itself; it spans ratings from .2 to .8 for V2 and .3 to .8 for V3 with error bars showing that every value occurred. The two boxplots for V2 and V3 are also very similar overall. However, this does not account for the generally low ratings on V1, which exhibit similar acoustic values as V2. One explanation for those low ratings on V1 in *namuta*-cases may then stem from the fact that listeners had to listen to their comparable word form with a long vowel as well (e.g. if one listened to *namuta* ‘cloud-acc’, they also had to listen to *naamu* ‘cloud’, the form with a long vowel). In those forms, listeners attended to the very long first vowel so that one could argue that this shifted the prominence ratings away from V1 in *namuta*-cases as listeners may have concluded that length could be a cue on V1, it is absent now, and the prominence needs to go on a different vowel, be it V2 or V3. The fact that the acoustics do not give a precise picture for anything that could cue prominence could then be reflected in the two boxplots in V2 and V3 in figure 14a, where one can see that the ratings between V2 and V3 are almost identical. This highlights that listeners were unsure where to locate prominence when being forced to pick a more prominent syllable.

Considering statements, the prominence seems to be located on V3 according to the listeners. The problem here is that none of the acoustic measurements are available for V3. However, a quick informal look at most tokens reveals that it seems that the pitch peak in *namuta*-cases very often lies on V3. Further supports come from the fact that the contour on V2 is often a rise sometimes with a pitch difference of as little as 3Hz, sometimes with a 30Hz difference. The pitch peak on V3 could then be one explanation for the pattern of the perception experiment[10] but I would suggest to remain cautious about those pitch peaks. The main reason is that the preceding segment is usually a voiceless stop /p/, /t/, or /k/, which could cause microprosody and therefore higher peaks, small peak excursions, or jumps. Furthermore, the third and final segment may already be part of a rise for a question, causing the peak to be on V3, too. A further possible reason for this peak on V3 could be potential continuation rises before intonational phrase boundaries or pauses inserted by the consultant. Finally, little is known about the intonational system of Hiaki in general. It could therefore be the case that some sentential information or particular intonational patterns cause some prominence to be located on the V3. One possible explanation here could be that the suffix added contains crucial information if one compares *mina-ta* and *mina-po*. In the latter case, the *suffix -po* indicates location translated as 'at' whereas it marks the object in the first case. Intonation or prominence could then be used to stress this suffix.

A problem arguing with a pitch peak on V3 is that this is based on a fairly small sample measured informally and most often, the tokens included in experiment 1 do not display a pitch tracker on V3; and if there is a pitch track, it may include several errors and jumps. Intensity peaks on V3 were therefore checked informally as well. An interesting finding is that apart from

[10] Note that Gussehoven (2004) states that pitch is a driving cue in the perception of prominence in English and this could be one reason here because the listeners were native English speakers.

all affixed version of *naawa* 'root', most tokens exhibits higher intensity values on V3 than on V1 or V2, up to 3db. This would then stand in line with the prominence ratings on V1 in *naamu*-cases, where intensity also peaked on the vowel with the highest prominence ratings. Yet, this may not be due to the intensity alone as the length may significantly contribute here as well but those sample measurements provide an interesting insight that it may be well worth to include V3 of all *namuta*-cases when analyzing pitch accent and prominence in Hiaki.

As a result, it remains difficult to find a clear answer why it is the third vowel in *namuta*-cases that was judged to be most prominent because precise acoustic measurements for V3 are not available, the differences between V1 and V2 in *namuta*-cases is minimal, no previous study has included prominence on the third vowel, and the factors just outlined may all contribute to why listeners rated V3 as most prominent. All this then means that the combination of both experiments could not reveal clear acoustic correlates of the morphologically determined pitch accent. While *naamu*-cases showed that length, intensity, and a low tone corresponded to listeners prominence ratings, no clear difference in exactly those measurements for *namuta*-cases emerged and listeners judged the vowels towards the end of the word as more prominent.

4.3. The potential shift of the morphologically determined pitch accent

The findings presented above cannot answer the second research question about a potential shift of the morphologically determined pitch accent. It is particularly the acoustic measurements on *namuta*-cases that simply show no differences between the first and second vowel so that it remains difficult to claim that there has been a pitch accent shift compared to *naamu*-cases. Surely, one could argue that while prominence is located on the first vowel in *naamu*-cases, listeners made it clear that the prominence is probably not located on the first vowel in *namuta*-cases. Yet, I would suggest that it is not as simple as the perception study implies. Looking at figure 14a for *namuta*-cases perceived in isolation, the difference between V1 and V2 is interesting in that V2 received higher ratings but exhibits similar acoustics as V1 and other things such as sentence intonation could have not contributed here as well because those are the tokens presented in isolation. Note though that as outlined earlier, the reason might be the absence of a long vowel so that listeners might have shown a tendency to not locate prominence on V1. Regarding statements, listeners marked prominence on V3 while V1 and V2 are rated similarly. Simply going by the listeners ratings, this would mean that prominence shifted from the first vowel in *naamu*-cases to the third vowel in *namuta*-cases, a pattern completely unattested in Hiaki and the jump from syllable one to syllable three seems generally quite far. The only possible explanation which would reduce the distance of the prominence jump could be that the prominence on the long vowel in *naamu*-cases attaches to the second mora of the word which then shifts to the third under affixation. It is not uncommon for languages to attach prominence as an autosegment (Hagberg, 1993; Demers et al., 1999) but this would need much more research in that the prominent would have to precisely identified to be on the second mora of the long vowel in *naamu*-cases followed by a careful investigation of V3 in affixed *namuta*-cases. This argument would also stand in direct contrast to Demers et al. (1999), who claim that the "overall position of the tone in the word remains constant: in this case on the second mora" (p. 46). Keeping the acoustics from experiment 1 in mind, I would suggest to treat such an interpretation with *extreme* caution until further data are available.

As a result, the study could not answer the question about a potential pitch accent shift because acoustic measurements on *namuta*-cases are close to be identical for V1 and V2 and measurements on V3 were not included. Therefore, it remains difficult to draw reasonable conclusions about a potential pitch accent shift since it also remains unclear where prominence in those *namuta*-cases is located. The perception experiment provided hints that listeners perceive more prominence on the second or third vowel but no acoustic correlate for this finding could be detected.

4.4. Vowel length as prominence marker?

A final thought on Hiaki prominence is concerned with vowel length. Vowel length is a decisive cue for prominence and also cues stress in English (Gussehoven, 2004). Cole et al. (2010) analyzed intensity and duration showing that the correlation between perceived prominence and duration is the strongest. Experiment 1 falls into this category as the very long first vowels in *naamu*-cases received extremely high prominence ratings in the RPT method. One could then argue that regardless of what other cues might have been present in *naamu*-cases, the length of the first vowel could have overridden all other prominence features because the vowels are so long. Looking at *namuta*-cases, the story may look different as the length of V1 is similar to the length of all other vowels in such cases so that other acoustic correlates may then have to cue prominence. Although it remains unclear which acoustic cues would fulfill this role, the perception task illustrates that listeners on the one hand are unsure whether prominence is located on V2 or V3 but they do agree in both isolation and statements that V1 is the least prominent vowel; possibly because they are biased due to the absence of the long V1 in *naamu*-cases. Consequently, one may conclude that vowel length serves as main cue for prominence in *naamu*-cases, overriding other prominence cues but those cues will then have to cue prominence in *namuta*-cases due to the lack of vowel length.

Another interesting thought comes here form Demers et al. (1999), who state that "vowel duration is a competing marker of prominence (perhaps analogous to stress clash) and in order to maximize the prominence of the first syllable bearing a high tone, a mora is dropped in adjacent syllables containing two moras" (p. 46). What the authors suggest is that if a high tone does not align with the bimoraic, i.e. long vowel, the long vowel will shorten if a high tone occurs to the right of such long vowels so that the prominence of that high tone will increase and does not have to compete against the vowel length in the preceding syllable. This argument is indeed sound as length is a powerful cue to prominence, as shown in experiment 2. They illustrate their claim with examples coming from tokens with a lexical pitch accent, e.g. *kaáte* VS *kakáate* 'are walking' and *káate* VS *kákate* 'is building a house' (Demers et al., 1999, p. 46). However, the first question is then why we can observe a pitch peak on the second vowel in *naamu*-cases whereas the first vowel is long and according to the statement above, it would have to shorten then; this is not the case. It would make sense though to follow along those lines for all *namuta*-cases because listeners perceived prominence on either V2 or V3 but never on the first vowel. Since those cases include a shorted first vowel, we would assume that prominence, or as Demers et al. (1999) argue a high tone, should occur on V2 but the acoustic correlates in experiment 1 do not confirm that. One has to note though that Demers et al.'s (1999) claim may be limited to particular cases such as reduplicated forms or words with a lexical pitch accent. As a result, the length of the first vowel in *naamu*-cases may have the power to override the prominence of a potential high tone, which could also shift due to certain morphologically processes. The long vowel would therefore be shortened so that the high tone not located on that long vowel does not have to compete against the prominence of that vowel. The shortening of long vowels thus enhances the prominence of a high tone. The only problem with this analysis is that the acoustics found in experiment 1 cannot confirm this assumption. We can only observe that the long vowels are indeed very prominent in both the acoustics and the perception.

5. Conclusion

Previous research investigating the morphologically determined pitch accent in Hiaki has resulted in different conclusions, not providing a clear picture about the acoustic correlates and where the pitch accent occurs, i.e. whether it attaches to morae or syllables. This study attempted to analyze potential acoustic correlates such as pitch peak or intensity peak and also addressed the pitch accent from a perceptual side with native English-speaking listeners. While tokens such as *naamu* ‘cloud’ showed extremely long vowels, pitch peak, and intensity peak differences making them partly align with previous research, it was predominantly the acoustic measurements on the affixed word forms such as *namuta* ‘cloud-acc’ which were almost identical on the first and second vowel. Those measurement could therefore not provide new understandings about the morphologically determined pitch accent, i.e. about the precise acoustic correlates and the accent's placement. The following perception study highlighted once more that the phonologically long vowels which are also long phonetically serve as massive prominence anchor as the great majority of listeners marked those vowels as prominent. Apart from that, the perception task has provided little new evidence about how the pitch accent could be realized and where it goes as listeners where unsure where to locate prominence in *namuta*-cases. As a consequence, further research is necessary to uncover the acoustic correlates in first place and then gain a deeper understanding of Hiaki phonology in terms of pitch accent placement.

Acknowledgements

First and foremost, I would like to thank Maria and Santos for their incredible help in eliciting all the stimuli. On a weekly basis, they would come up with new sentence of the same words over and over but they would always patiently help us out with new creations. Without the help of Tim Mahrt, the perception experiment would still not run properly. He answered all my questions about LMEDS and tried to fix every problem I encountered. Last, I would like to thank my committee members Natasha Warner and Miquel Simonet for their awesome input with regards to the experiments and possible solutions as well as Heidi Harley for reading a phonetics paper and mentoring me to the point where I could write this paper.

References

Alvord, S. (2006). *Spanish intonation in contact: The case of Miami Cuban bilinguals* (Doctoral Dissertation). Available at https://server15999.contentdm.oclc.org/cgibin/showfile.exe?CISOROOT=/IR&CISOPTR=1055

Barnes, D. (2013). Age-related changes to the production of linguistic prosody (Master's Thesis). *Open Access Theses*. Paper 17. Available at https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1016&context=open_access_theses

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* [Computer Program]. Version 6.0.36. Retrieved November 11, 2017 from http://www.praat.org/

Bogomolets, K. (2014). Acoustic Correlates of Arapaho Prominence. *Linguistics Graduate Theses & Dissertations*. Retrieved October 17, 2017 from https://scholar.colorado.edu/ling_gradetds/30

Bruce, G. (2005). Intonational prominence in varieties of Swedish revisited. In Sun-Ah Jun (Ed.), Prosodic typology: The phonology of intonation and phrasing (pp. 410-429). Oxford: Oxford University Press.

Bullock, Barbara E. 2009. Prosody in contact in French: A case study from heritage variety in the USA. *International Journal of Bilingualism*, *13*(2), 165-194. doi:10.1177/1367006909339817

Cole, J., Mo, Y., Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425-452. doi: 10.1515/labphon.2010.022

Cole, J., Mahrt, T., & Hualde, J. (2014). Listening for sound, listening for meaning: Task effects on prosodic transcription. *Proceedings of the International Conference on Speech Prosody*, *7*, 859-863. Retrieved from https://pdfs.semanticscholar.org/2dcc/ef9fa644ff27d561c8dca5928274c5ea1e8a.pdf

Cole, J., Mahrt, T., & Roy, J. (2017). Crowd-sourcing prosodic annotation. *Computer Speech & Language*, *45*, 300-325. doi:10.1016/j.csl.2017.02.008

Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1), 1-19. doi:10.5334/labphon.29

Crumrine, L. (1961). *The phonology of Arizona Yaqui with texts*. Tucson, AZ: The University of Arizona Press.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustic Society of America*, 105(3), 1877-1888. doi:10.1121/1.426724

De Leeuw, E. (2009). *When your native language sounds foreign: A phonetic investigation into first language attrition* (Doctoral Dissertation). Available at http://etheses.qmu.ac.uk/119/1/119.pdf

De Leeuw, E., Menne, I., & Scobbie, J. (2011). Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism*, *16*(1), 101-116. doi:10.1177/1367006911405576

De Leeuw, E., Tusha, A., & Schmid, M. (2017). Individual phonological attrition in Albanian-English late bilinguals. *Bilingualism: Language and Cognition*, *21*(2), 1-18. doi:10.1017/S1366728917000025

Dedrick, J., & Casad, E. (1999). *Sonora Hiaki language structure*. Tucson, AZ: The University of Arizona Press.

Demers, R., Escalante, F., & Jelinek, E. (1999). Prominence in Hiaki words. *International Journal of American Linguistics*, *65*(1), 44-55. doi:10.1086/466375

Dupis, K., & Pichora-Fuller, M. (2010). Use of affective prosody by young and older adults. *Psychology of Aging*, 25(1), 16-29. doi:10.1037/a0018777

Fabián, C., & Duarte, A. (2006). *Yaqui: A double-aligned trochaic feet language*. Retrieved March 17, 2018, from http://arizonahiaki.org/wp-content/uploads/2016/08/Yaqui_Pitch_Constantino.pdf

Faraclas, N. (1984). Rivers Pidgin English: Tone, stress, or pitch-accent language. *Studies in the Linguistic Sciences, 14*(2), 67-76. doi:10.4324/9780203192801

Fraenkel, G. (1959). Yaqui phonemics. *Anthropological Linguistics*, 1(5), 7-18. Retrieved from http://www.jstor.org/stable/30022203?seq=1#page_scan_tab_contents

Gordon, M. (2014). Disentangling stress and pitch-accent: A typology of prominence at different prosodic levels. In H. van der Hulst (Ed.), *Word Accent: Theoretical and Typological Issues* (pp. 83-118). Cambridge: Cambridge University Press.

Gussehoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hagberg, L. (1993). *An autosegmental theory of stress* (Doctoral Dissertation). Available at ProQuest Dissertations and Theses. (UMI No. 9328564)

Hagberg, L. (2008, October 3). *An acoustic analysis of Yaqui stress*. Presentation at the Friends of Uto-Aztecan Conference (FUAC), Tucson, AZ.

Haraguchi, S. (1988). Pitch accent and intonation in Japanese. In H. Van der Hulst, & N. Smith (Eds.), *Autosegmental studies on pitch* (pp. 123-150). Dordrecht: Foris Publications.

Hyman, L. (2009). How (not) to do phonological typology: The case of pitch-accent. *Language Sciences*, 31(2-3), 213-238. doi:10.1016/j.langsci.2008.12.007

Inkelas, S., & Zec, D. (1988). Serbio-Croatian pitch accent: The interaction of tone, stress, and intonation. *Language*, 64(2), 227-248. doi:10.2307/415433

Ladd, D. (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? Italian Journal of Linguistics, 18(1), 19-38. Retrieved from http://www.italian-journal-linguistics.com/wp-content/uploads/2.Ladd_.pdf

Ladd, D., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, *107*(5), 2685-2696. doi:*10.1121/1.428654*

Mahrt, T. (2016a). *LMEDS: Language markup and experimental design software* [Computer program]. Version 2.5. Retrieved January 18, 2018 from https://github.com/timmahrt/LMEDS

Mahrt, T. (2016b). LMEDS: A platform for collecting prosodic annotation online [PowerPoint slides]. Available at https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source

=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjbsfv-8PnZAhVP4GMKHUxAAdkQFggnMAA&url=http%3A%2F%2Fwww.timmahrt.com%2Facademic_files%2Fmahrt_lmeds_cologne_2016_04_25.pptx&usg=AOvVaw2X0CWzyRTHsIEPrwJU7dpb

Mo, Y., Cole, J., & Lee, E. (2008). Naïve listeners' prominence and boundary perception: In P. Barbosa, S. Madureira, & C. Reis (Eds.), Proceedings of Speech Prosody 2008 (pp. 735-738). Available at http://www.isca-speech.org/archive/sp2008/papers/sp08_735.pdf

Mona, F., Valenzuela, H., & Shaul, D. (1999). *Hippocrene standard dictionary Yoeme-English English Yoeme: With a written comprehensive grammar of Yoeme language*. New York: Hippocrene Books.

Mayr, R., Price, S., & Mennen, I. (2012). First language attrition in the speech of Dutch-English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition*, *15*(4), 687-700. doi:10.1017/S136672891100071X

McGlone, R., & Hollien, H. (1963). Vocal pitch characteristics of aged women. *Journal of Speech, Language, and Hearing Research,* 6, 164-170. doi:10.1044/jshr.0602.164

Pitt, M.A., Dilley, K., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release). Columbus, OH: Department of Psychology, The Ohio State University.

Ptacek, P., Sander, E., Maloney, W., & Jackson, D. (1966). Phonatory and related changed with advanced age. *Journal of Speech, Language, and Hearing Research*, 9, 353-360. doi:10.1044/jshr.0903.353

Printer, G., Mizuguchi, S., & Tateishi, K. (2014). Perception of prosodic prominence by L1 and L2 speakers of English. *Proceedings of InterSpeech 2014*, 544-547.

Rafat, Y., Mohaghegh, M., & Stevenson, R. (2017). Geminate attrition across three generations of Farsi-English bilinguals living in Canada: An acoustic study. *Ilha do Desterro*, *70*(3), 151-168. doi:10.5007/2175-8026.2017v70n3p15

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody transcription. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1), 1–36. doi:10.5334/labphon.108

Simons, G., & Fennig, C. (2017). *Ethnologue: Languages of the world*. Dallas, TX: SIL International.

Stathopoulous, E., Hubher, H., & Sussman, J. (2011). Changes in acoustic characteristics of the voice across the life-span: Measures from 4-93 year olds. *Journal of Speech, Language, and Hearing Research*, 54, 1011-1021. doi:10.1044/1092-4388(2010/10-0036)