1. Introduction

Though the Chatino languages of Oaxaca have been studied for some time now, a rigorous acoustic analysis of the tonology of the language family has not yet been undertaken, and the mechanisms by which its speakers distinguish among the many phonemic tones are not understood. With the aim of eventually performing a battery of perceptual experiments to test several potential acoustic cues for their salience or irrelevance for a speaker’s ability to perceive tone, an initial series of data-seeking measurements must be taken to understand the range of empirical variation in high-contrast speech to develop appropriate stimuli for probing the acoustic cues used by hearers of the language.

2. Background

2.1 The Chatino Family

Chatino is a shallow language family of the Mexican state of Oaxaca. The Chatino family is coordinate with the Zapotec languages within the Otomanguean language family.
Various proposals for divisions of languages within the Chatino family have been proposed, from three varieties as proposed by Franz Boas (1913), to the six varieties proposed by Ethnologue. The chief variation in these classification is the division or cohesion between the varieties of what is either considered one variety (Boas’s “First Dialect”, CDLP’s Eastern Chatino) or as many as four (Ethnologue’s Zacatepec, Nopala, Western Highland, and Eastern Highland). The variety to be studied in this proposal is that of the communities of San Juan Quiahije and Cieneguilla, a variety (henceforth SJQ) of Eastern Chatino—Western Highland Chatino according to the Ethnologue classification (Gordon 2005).

2.2 Relevant Phonology and Syllable Structure

SJQ Chatino is a highly innovative (both segmentally and tonally) monosyllabic variety of Eastern Chatino, whose maximal syllable structure is CCVC, permitting many onset clusters (as well as prenasalization, represented orthographically with n- or m- preceding another obstruent) and only one coda consonant (the glottal stop, represented orthographically by ʔ).

Vowels exhibit a nasality contrast (which phonetically affects vowel quality), with nasalization represented by an orthographic -n following the vowel (Rasch 2002).

2.3 Relevant Tonology

2.3.1 Lexical Tones of SJQ

Among the simplex tones of SJQ\(^1\) are three rising contour tones 20, 42 and 40\(^2\); three falling contour tones 14, 24, and 04; eight linear tones (which may be steady rises or falls) 0, 10, 1, 2, 3, 32, weak 4 and strong 4; and one falling-rising tone 140. Complicating the matter of tone in SJQ are the presence of floating tones and the processes of sandhi which convert some tones

---

\(^1\) Certain processes such as contraction and especially the first-person inclusive conjugation of verbs produce long vowels with what appear to be a sequence of two dissimilar tones produced in full (such as the 24-2 sequence in ykween\(^2\)̄  we (inclusive) said’), and are not to be discussed here.

\(^2\) Tone is indicated by superscript numbers, with larger numbers representing lower tones. SJQ Chatino is described with five tone levels, 0-4, with 0 representing a rather high tone.
into different tones in particular contexts. Certain tones such as 14, 10, and 40 are said to possess a floating tone (itself represented by 0) which only surfaces when a word carrying such a tone is followed by a possible landing site, namely tones 1, 2, and some versions of tone 4 (these tones are divided into ‘strong’ and ‘weak’ categories), 24, and 14 (Cruz and Woodbury 2006).

2.3.2 Tone Sandhi

Tone sandhi chiefly affects (and serves to distinguish) the strong and weak versions of tone 4. Strong versions of tone 4 only undergo a change when preceded by another tone 4, in which case the second 4 becomes a rising tone 32. Weak 4 tones surface as 24 after tone 1, 3, and 20; tone 32 after tone 24; tone 1 after tone 2; and tone 0 after any syllable bearing a floating tone (Cruz and Woodbury 2006).

3. Significance

A study of the perceptual cues used by speakers of SJQ Chatino has the ability to inform several fields of research: Otomanguean and general tone perception, Chatino and Zapotecan historical linguistics, as well as comparative descriptions of Eastern Chatino varieties. Given the apparent division between types of tones in SJQ Chatino (contour and linear), it could be of interest to tone perception literature if different mechanisms appear to distinguish between different kinds of tones. Also, since all Otomanguean languages are tonal it stands to reason that proto-Otomanguean might have had some form of tonal distinction. If so, then the varying types of tonal systems must have been inherited from a common ancestor and a better understanding of the processes of tone in one variety cold lead to a better reconstruction of the family, as well as providing a better picture of how tonal systems might change over time. Finally, knowledge of how tone is perceived in one variety of Eastern Chatino might (or more interestingly might not) be able to explain how tone is perceived in other varieties, many of which report varying levels
of intelligibility of SJQ speech, helping to support the cohesion of the language or its division into any number of smaller languages.

4. Methodology

Tokens of each simplex tone were recorded over five repetitions each of three words containing the tone. Tokens were read in isolation by a female native speaker using a self-paced word list wherein tokens were written in the practical orthography (including numerical tone indications) accompanied by a gloss in Spanish. Tokens were selected to be common monomorphemic words in the SJQ variety of Eastern Chatino. To facilitate the easy comparison of tone, words containing phonemic vowel nasalization (which would phonetically lower f0) or a glottal stop (which may affect f0) were not chosen for recording when possible. Vowel duration was measured by hand using the PRAAT acoustic analysis software (Boersma and Weenik 2009), and from these measurements, measurements of f0 at 20% 50% and 80% of the vowel’s duration were collected by an automated script.

5. Results of the Measurements

From this data, duration-normalized pitch tracts of each tone were created by averaging the f0 value for each tone’s tokens at each of the three measurement points and plotting these points on a graph (Figure 1). For ease of legibility, the information in this graph is reproduced below, divided into contour (Figure 2) and linear tones (Figure 3).

---

3 The speaker in question, like many Chatino speakers, is bilingual with Spanish.
4 Given the fact that tone 40 only occurs in first-person singular verb conjugations, it co-occurs with nasalization. Tone 04 also tends to occur in syllables closed by the glottal stop.
Figure 1

Mean f0 in Hz (along the y-axis) for each tone produced by one speaker at 20%, 50%, and 80% of vowel duration (along the x-axis).

Figure 2 shows the duration-normalized pitch tracts of the contour tones recorded by one speaker. Tones 14 and 24 appear to have the same or similar pitch tracts, and in fact, this speaker has difficulty discerning between these syllables in isolation. Tone 140 also does not have the particularly low midpoint which its tone number would suggest.
Figure 2.

Mean f0 in Hz (along the y-axis) of contour tones produced by one speaker at 20%, 50%, and 80% of vowel duration (along the x-axis).

Figure 3 shows the mean pitch tracts of the linear tones of the speaker. The values produced for tone 10 (the second highest pitch on the chart) are suspect, as are the values as the speaker attempted to produce a tone 10—which is generally only found in contexts where syllable of tone 1 with a floating tone is followed by a tone which does not allow for the appearance of the floating tone—in isolation. Tone 10 is a high rising tone with slight elbow, which is not in evidence here, and pending identification of these tones as 10 in isolation by speakers, is not to be considered in the present analysis.
6. A Potential Tone Perception Experiment

6.1 A Possible Methodology

Taking the highest pitch in the data (465 Hz) and the lowest (165 Hz) as the ceiling and floor of the pitch range, the pitch of an utterance may be altered at any point in its duration, producing tokens whose pitch may rise, fall, or rise then fall over the duration of the syllable. The token which will be used for this task will be kla² ‘pool’ since its pitch track is generally level and falls within the middle of the pitch space. The choice of kla² is doubly apt since a great number of minimal pairs may be constructed from this string of segments as is shown in Table 1. The only tones not attested with this string of segments are 10 and 40 (which would only be possible in a nasal context).

Figure 3: Mean f0 in Hz (along the y-axis) for each linear tone produced by one speaker measured at 20%, 50%, and 80% of vowel duration (along the x-axis).
Table 1:

<table>
<thead>
<tr>
<th>Lexeme</th>
<th>Gloss</th>
<th>May surface as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>kla¹</td>
<td>loom</td>
<td></td>
</tr>
<tr>
<td>kla²</td>
<td>pool</td>
<td>kla⁰, kla²⁴</td>
</tr>
<tr>
<td>kla³</td>
<td>dream</td>
<td>kla³²</td>
</tr>
<tr>
<td>kla⁴d</td>
<td>POT.be.born</td>
<td>kla¹⁴⁰, kla¹⁴⁰⁰</td>
</tr>
<tr>
<td>kla⁴f</td>
<td>corn plant</td>
<td>kla⁰⁴</td>
</tr>
<tr>
<td>kla⁵</td>
<td>POT.release</td>
<td></td>
</tr>
<tr>
<td>kla¹⁴</td>
<td>POT.dance</td>
<td>kla¹⁴⁰</td>
</tr>
<tr>
<td>kla²⁴</td>
<td>twenty</td>
<td>kla⁰⁴</td>
</tr>
<tr>
<td>kla⁴²</td>
<td>POT.arrive.2s</td>
<td></td>
</tr>
</tbody>
</table>

Attested citation forms and sandhi-produced forms of kla.

Given the ample range of the pitch as well as the relatively compact pitch space occupied by all tones except tone 4—only 16 Hz appear to distinguish the mean f0 values of tones 2 and 3—an exhaustive, fine-grained analysis of these tones would quickly exhaust any listener’s patience. To make the task more manageable, perception of linear and contour tones may be separated, and the lowest reaches of the pitch space can be eliminated for linear tones, removing from consideration the pitches found only in utterances of tone 4. This would raise the pitch floor to 298 Hz for linear tones and lower the ceiling to 441 Hz, leaving 143 Hz of space in which the pitch tract may be altered with greater precision without producing a prohibitively large number of tokens for listeners to distinguish.

From 445 Hz, pitch may be decreased in 25 Hz increments until 295 Hz, creating seven possible levels of pitch for each of the three time points within the vowel, creating 343 potential words, which would be played for speakers who would then indicate what word was heard. Included in these stimuli would be the tokens produced for contour tone perception, whose pitch floor and ceiling would be 145 Hz and 495 Hz, with pitch targets at 50 Hz increments, producing

---

⁵ POT refers to the potential aspect, one of the four aspects in which a Chatino verb may appear.
another 343 potential words (several of which would overlap with the stimuli produced for linear tones). Figure 4 shows the various pitch targets for manipulation, with some sample pitch tracts drawn.

Figure 4

![Pitch targets in Hz for the three points along the vowel duration with some sample pitch tracts drawn.]

6.2 Expected Results

Based on the measurements given here, it is expected that speakers will be incapable of distinguishing natural utterances of weak and strong tone 4 and tones 14 and 24 based on pitch tract data alone; any ability to do so would suggest that another feature is being used to distinguish between each pair, and further investigation would be necessary. For the perception of manipulated tokens, it is predicted that speakers will use absolute pitch height and the value of the slope of the pitch (roughly positive or negative) to determine the identity of linear tones, that is, pitch slope over the entire syllable and mean pitch of the entire syllable is predicted to interact with tone identity. For contour tones, it is predicted that speakers would base their ability to identify tones based on the magnitude of the differences between midpoint and endpoint pitch, as

---

6 Other native speakers claim to produce distinct tones 14 and 24 (Emiliana Cruz p.c.).
well as the direction of the pitch movement between these points, i.e. magnitude and direction of
the pitch slope between these points is expected to predict the speaker’s ability to identify these
tones.

At the same time, it is predicted that many of the tokens will be rejected or described as
having no tone, especially in cases where the beginning pitch and the midpoint pitch are greatly
mismatched. Since SJQ tones appear to remain on their native syllables, it is not expected that
speakers will accept say, a pitch tract from 495 Hz to 195 Hz to 195 Hz (a very sharp fall
followed by a very low plateau), as a tone 4, ignoring the initial pitch target, but will reject these
tokens as not being Chatino, though a robust identification of such a pitch tract as the low tone 4
may indicate that the beginning pitch of a vowel is less salient for tone perception.

6.3 Limitations

Some limitations on this study include its reliance on one speaker’s production of tones to
classify and produce stimuli for other speakers to distinguish, tacitly suggesting that the
performance of this one speaker will generalize across all speakers of SJQ Chatino, which, while
a fairly standard assumption, still has the potential to confound the experiment. For example, the
speaker’s very large pitch range may initially be a source of confusion for listeners: the speaker’s
mid tones may be initially perceived as high tones unless some short familiarization with the
speaker’s voice is allowed. Also, the presentation of tokens in isolation may cause listeners to be
reluctant to identify tokens with sandhi-produced tones, since the contexts which motivate the
appearance of those tones would be lacking, unless speakers were told these forms were excised
from sentences, which would introduce an undesired unnaturalness to the stimuli. Similarly,
even presentation of stimuli spliced into carrier sentences would prime listeners to not hear
certain tones which could not surface in some contexts. Even these confounding effects on the
experiment would open new avenues of research since they would suggest that sandhi processes are opaque and instead of perhaps aiding word recognition or speeding lexical access, inappropriate context-less sandhi may block lexical access entirely.
References


