Pitch and phonological perception of tone in the Suruí language of Rondônia (Brazil): identification task of LHL and LHH tonal patterns

Julien Meyer

1Área de Linguística, Museu Goeldi, Ministério de Ciência, Tecnologia e Inovação, Belém, Pará, Brasil
2Sound Communication and environmental auditory Perception Research Group, Paris, France

julmeyer.lab@gmail.com

Abstract

The present study explored the relation between pitch and phonological perception of tone in the Suruí language of Rondônia (Brazil), given that pitch realization of tone in Suruí is a complex phenomenon. F0 values and pitch contour of the vowel nucleus were found to influence the perception of tone in this language of the Mondé family, but these effects were found to be dominated by the influence of the pitch contour of the preceding syllable. These perceptual results, completed by an analysis on production, contribute to indicate the existence of a Mid tone in Suruí, allophone of the High tone, which would be the result of a regressive downstep.

Index Terms: Tone perception, Suruí language, Mondé languages, AXB test, downstep

1. Introduction

The Suruí language of Rondônia (auto denomination Paiteer) belongs to the Mondé family, one of the five families of the Tupi stock which are found in the state of Rondônia in Western Brazil. The Suruí began to be in continuous contact with Portuguese at the end of the 60s. The current population is of more than 1000 persons living in the Indigenous Territory Sete de Setembro, in Rondônia; they all speak their native language. The Mondé family is composed of two other languages: Salamây (Mondê) and a language composed of four dialects: Gavião of Rondônia, Zoró, Cinta Larga, and Aruá. All of these languages show distinctive tone and syllable length [1]. The Suruí tone system is relatively simple with two phonological level tones, High (H) and Low (L) and the tone bearing unit is the vowel, with only one tone per vowel, no matter its length [2].

Despite the existence of this simple two level tone system in Suruí, its pitch realization is a complex phenomenon. Guerra [2] first observed a relatively simple phonetic pattern for monosyllables, in which the low tone is manifested as a falling pitch and the high tone as a constant or rising pitch. Looking at disyllables, she noticed that the relation between tone and pitch is more complex; it depends on the position the syllable occupies inside the word and on the tone of the surrounding syllables. At the end of a disyllabic word a high tone tends to be constant or only rises a little bit, whereas at the beginning of such words it tends to be constant when followed by a high tone and to rise when followed by a low tone. Concerning low tones, they always fall at the end of disyllabic words whereas they were described by Guerra as either constant or falling when followed by a low tone and falling when followed by a high tone [2]. Therefore, in the light of the existing literature, one of the principal subtleties of the relation between pitch and the phonological perception of tone appears to be that the contour of the curve (i.e. constant, rising, falling) is highly important for the perception, and, according to Guerra, it is perhaps more important than the pitch value itself.

To extend the research on the relation between phonetic pitch and phonological perception in the Suruí tonal system, this investigation explored here one particular aspect of LHL and LHH tonal profiles. It focused on two three-syllable utterances phonologically distinct only by the tone on the last syllable but at the same time phonetically distinct by the pitch on the two last syllables. Among the possibilities of utterances that complied with these conditions, the tested pair was [i.a.o] [kap] (caiman fat) / [i.a.ô] [káp] (caiman egg). A perceptual identification experiment was conceived: first, a series of artificial target words were built by modifying the pitch contours of the words [i.a.o] [kap] and [i.a.ô] [káp], which were considered as prototypes.

Next, the study consisted in testing if native speakers identified each of the target words either with [i.a.o] [kap] or with [i.a.ô] [káp] by presenting stimuli to Suruí listeners in an AXB identification task (2AFC), following some aspects of the methods used in Hallé et al [3].

2. Method

2.1. Participants

The tested participants were 8 native Suruí speakers of 19 to 35 years old who live in the villages of Lapetanha and Pin Paite on the federally protected Indigenous Territory called ‘Terra Indígena Sete de Setembro’. All participants reported normal hearing. They were bilingual Suruí-Portuguese, literate in Portuguese and semi literate in Suruí as no standardized written system of Suruí language exists for the moment [4]. Three other Suruí speakers participated to the conception and the recording of the stimuli. All participants provided oral (recorded) informed consent to the study, which was conducted in accordance with the guidelines of the Declaration of Helsinki and of the Museu Paraense Emílio Goeldi. Moreover the data were processed anonymously.

2.2. Materials

2.2.1. Reference utterances of the experiment

The materials used in the experiment reported below were derived from natural utterances. To prepare the stimuli, 5 repetitions of each [i.a.o] [kap] and [i.a.ô] [káp] utterance were
recorded with three different male speakers in a calm and non reverberant room of the indigenous nongovernmental association Metareilá (cf. acknowledgements). Pitch contours were observed on each syllable. For both types of utterances, the first low tone was characterized by a short constant pitch followed by a rapidly falling pitch (-14 Hz in 0.1 sec. in Figure 1 and -12 Hz on Figure 2). However, the pitch contours differed on the next syllables. Namely, for each utterance of [βa.ó] [kap] or [βa.ó] [kap], the pitch contour of the second syllable depended on the nature of the third tone. When the third tone was low ([kap] marked by a falling pitch of 14 Hz and 20 Hz in Figure 1 and 2 (left)), the pitch of the second syllable /ó/ rose rapidly (+14 Hz and +17 Hz in Figures 1 and 2, left) and finished constant. On the other hand, when the third tone was high ([kap] marked by a short constant step at the end on the vowel nucleus, like in Figures 1 and 2, right), then the pitch of [ó] was constant.

![Figure 1: Spectrogram and pitch contour (in blue) of the two selected utterances [βa.ó] [kap] (left) and [βa.ó] [kap] (right).](image1)

![Figure 2: Spectrogram and pitch contour (in blue) of [βa.ó] [kap] (left) and [βa.ó] [kap] (right) by a second speaker.](image2)

For the present experiment, the clearest productions of the speaker who spoke the most distinctively were chosen (that is, those of Figure 1).

### 2.2.2. Other Suruí words with similar pitch profiles

The particular pitch behavior just described for [βa.ó] [kap] and [βa.ó] [kap] is shared by several Suruí words with similar LHH and LHL tonal profiles. Of course, some differences appeared between words, sometimes due to consonant types that separate the vowels (particularly when a stop broke the continuity of f0 between the first and the second syllables, like in Figure 3, left), or sometimes due to a higher relative pitch on the second syllable in comparison to the first syllable (see for example the word [t∫irigá], Figure 3, right). However, comparisons with such words highlighted the important common aspects of all observed examples: (i) first, the second syllable was always characterized by a constant pitch contour in LHH words and a rising pitch contour in LHL words. (ii) Next, the second H tone of LHH utterances was always characterized by a lower pitch level than the one of the third H tone, as if the second one was ‘downstepped’. Interestingly, a constant pitch on the second syllable was sufficient to categorize its phonological tone as a high tone instead of a low tone, despite its relatively low pitch value. As noted by Guerra [2, p.77] a low tone in Suruí would have been characterized by a falling pitch at the end of a word or before a high tone (see a LHL word on Figure 4).

![Figure 3: Variants of the possible pitch contours of LHH tonal patterns ([ladájip] means ‘larva-fat’, [t∫irigá] means ‘to tie’), by a third speaker.](image3)

![Figure 4: Typical pitch contour of LLH tonal patterns. Low tones of the two first syllables are indicated by a rapidly falling pitch ([akok∫] means ‘firefly’).](image4)

### 2.2.3. Stimuli of the experiment

The stimuli presented to the participants consisted of 14 different audio wav files, each containing three words: the two [βa.ó] [kap] and [βa.ó] [kap] prototypes, as well as a resynthesized target word built by modifying the fundamental frequency (f0) of the last two syllables of the prototypes. The audio tracks began with one of the two prototypic words (A) followed by the target (X), itself followed by the other prototypic word (B). Each word was preceded and followed by a silence of 500 ms. Therefore, each audio track had the following structure: silence-A-silence-X-silence-B-silence. Seven different artificial target words (Xa, Xb, Xc,Xd,Xe,Xf,Xg) were built and each target was included in two different audio tracks: one beginning by the prototype [βa.ó] [kap] and finishing by the prototype [βa.ó] [kap], the other beginning by [βa.ó] [kap] and finishing by [βa.ó] [kap].

The f0 contours of the seven different artificial target words of this experiment are here presented in schematic form (Table 1), together with the f0 contours of the two natural prototypes of Figure 1 (P1 and P2 in Table 1). The pitch patterns of the targets were carefully selected to test the impact on tone identification of various pitch configurations of the two last syllables /ó kap/. The f0 values of the target words were chosen within the range
of values naturally produced by the selected speaker (basically from 117 Hz to 138 Hz, see Figure 1). Only one of the tested targets broke this rule (Xg), and it was chosen to check the reaction of the listeners in a more artificial case. Technically, one of the two prototypic forms (P2) was chosen to serve as a base for building these stimuli. Next, using Praat software [5], the f0 was extracted, modified and resynthesized to build the 7 target words and also to rebuild the second prototype (P1). Thus, the possibility that other factors linked to duration, amplitude, or formant structure would influence the answers of the listeners was neutralized. To build the target words, The two possible pitch contours of the last syllable have been sometimes shifted up or down in frequency and sometimes also combined with the two possible pitch patterns of the second syllable [6]. More precisely, the pitch contour of the last syllable of P2 was either kept equal or modified into the pitch pattern of the last syllable of P1. When the selected pitch contour of the last syllable was the one of P2 ([kap]), it was sometimes shifted up/down of 7/14 Hz and combined with the tonal pitches of the preceding syllables of P1 ([ja.ó] [kap]) or of P2 ([ja.ó] [káp]). Moreover, when the pitch contour was the one of [kap] (P1), it was either shifted up of 21 Hz (Xg), or kept at the same pitch but combined with the alternative pitch pattern of the second syllable [6] (Xa). For resynthesis of the f0 curve, intonational landmarks such as turning points were used. After resynthesis, the new audio samples were tested on two extra participants for validating the naturalness of the pronunciation. That is, there was no "nonword" in the materials.

Table 1. List of pitch contours (prototypes and targets)

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Pitch Contour</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>[βa .ó] [kap]</td>
</tr>
<tr>
<td>P2</td>
<td>[βa .ó] [káp]</td>
</tr>
<tr>
<td>Xa</td>
<td></td>
</tr>
<tr>
<td>Xb</td>
<td></td>
</tr>
<tr>
<td>Xc</td>
<td></td>
</tr>
<tr>
<td>Xd</td>
<td></td>
</tr>
<tr>
<td>Xe</td>
<td></td>
</tr>
<tr>
<td>Xf</td>
<td></td>
</tr>
<tr>
<td>Xg</td>
<td></td>
</tr>
</tbody>
</table>

2.3. Design and procedure

As said in introduction, Suruí listeners participated in an AXB identification task (2AFC). Therefore, after listening to an audio track their task was to identify the target word X either to [ja.ó] [kap] or to [ja.ó] [káp] prototypes. This task differed from traditional identification tasks as listeners heard tone category prototypes surrounding the target stimulus. Such an option was chosen to ease the development of the experiment in a population unfamiliar to such perceptual tests. Participants were asked to speak out loud the sequence of the three words in Suruí language and then to translate this sequence in Portuguese for verification. The experimenter wrote down the answers and then played the next audiotrack.

A list of 14 audio tracks was played twice in a predetermined order to each subject. Here is the list: 1) P1XeP2, 2) P1XfP2, 3) P2XbP1, 4) P1XcP2, 5) P2XdP1, 6) P2XgP1, 7) P1XaP2, 8) P2XeP1, 9) P2XP1, 10) P1XbP2, 11) P2XcP1, 12) P1XdP2, 13) P1XgP2, 14) P2XaP1). Consequently, each target word (Xi) appeared four times for each person in the test.

3. Results

The results of the experiment are presented in the Table 2 below.

Table 2. Last tone perception as a function of target (average % for H and L)

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Pitch Contour</th>
<th>Last tone perception [average %]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>Xa</td>
<td></td>
<td>96.875</td>
</tr>
<tr>
<td>Xb</td>
<td></td>
<td>93.75</td>
</tr>
<tr>
<td>Xc</td>
<td></td>
<td>81.25</td>
</tr>
<tr>
<td>Xd</td>
<td></td>
<td>40.625</td>
</tr>
<tr>
<td>Xe</td>
<td></td>
<td>6.25</td>
</tr>
<tr>
<td>Xf</td>
<td></td>
<td>3.125</td>
</tr>
<tr>
<td>Xg</td>
<td></td>
<td>15.625</td>
</tr>
</tbody>
</table>

These results can be divided in three main groups as a function of the answers on the last tone. A first group constituted of Xa, Xb, Xc, in which the last tone was perceived as high (H). These have also in common to show a second syllable with the pitch pattern of P2, [ja.ó] [káp] (where the pitch of [ó] is mostly constant; like in Figure 1, right). A second group is constituted of Xe, Xf, Xg, in which the last tone was perceived as low (L).
These have in common to show a second syllable with the pitch pattern of P1 (rising rapidly). The last group is constituted of Xd alone, in which the results were much more distributed between high (40.625\%) and low (59.375\%). Moreover, for Xd the distribution of answers was asymmetrical as a function of the order of the prototypes that surrounded the target. Such a particular distribution of answers never occurred for the other targets. In this case, when the order of presentation was P2XdP1, Xd was identified in 87.5\% as [já.ó] [kap] (H tone on the last syllable), whereas it was identified in 68.75\% as [já.ó] [kap] (L tone on the last syllable) when the order of presentation was P1XdP2. On the contrary, for the other targets the difference found between the two orders of presentation was not significant statistically because it was either of one answer only (for Xa, Xe, Xf, Xg) or of none (for Xb, Xc).

4. Discussion and conclusions

This study explored a particular aspect of the relation between pitch contour and tone in Suruí. Therefore, the identification performance of the participants showed some interesting properties of Suruí tone production and perception. The first of these properties was that the pitch pattern of the second tone of the tested utterances strongly influenced the tone perception of the third syllable. This was typically showed by the emergence of two groups ((Xa, Xb, Xc) and (Xe, Xf, Xg)). These two groups were characterized by a correlation between the pitch contour of the second syllable and the tone judgment of the participants on the last syllable (H or L). Such a behavior of the participants was a perceptual confirmation of the importance of the phonetic f0 variation typically found in production between second syllables of LHH and LHL utterances (that is, constant vs. rising, cf. 2.2). As a consequence, we observed here both the perceptual and the productive importance in Suruí of the lowering of the first High tone in the case of two consecutive High tones in a word. To go deeper in the linguistic interpretation of this fact, the present study showed that the low constant pitch profiles of second syllables in LHH utterances presented the characteristics of a kind of Mid-tone, which would be an allophone of High, as it is in the Gavião language of Rondônia (that is, a H downstep that becomes M [7]). The difference with the sister Gavião language is that the downstep is here regressive instead of progressive.

Another aspect of tone perception highlighted by the present study concerned the influence of the pitch value of the proper syllable. For example, the fact that the listeners were generally hesitating in the case of Xd suggested that the pitch value was also important to Suruí listeners. Indeed, in this case, even if the pitch contour of the second tone was rising (typically announcing a low tone on the next syllable) the high pitch value of the following syllable pushed several listeners to judge that they had heard a high tone. Another proof of the influence of the pitch value of the proper syllable on the perception of its tone is provided by comparing the answers on Xb and Xc. In these two cases, the non rising second syllable pitch announced a following high tone, but when the pitch of this following tone was lowered of 14 Hz, the number of perceived low tone was tripled (even if it stayed below 20\%). Therefore, the difference between Xb and Xc showed that the pitch value influenced the perception. Such a conclusion was also confirmed by comparing the differences between Xd, Xe and Xf, in which the higher the pitch, the more the participants perceived high tones.

A last aspect on tone perception showed by this study concerns the role of falling pitches. First, the pattern of answers on Xa showed that a rapidly falling pitch contour - described by Guerra as systematically associated in Suruí to a low tone [2] - was not sufficient for the listeners to perceive a low tone if the preceding tone was a Mid one announcing a H tone. However, the comparison of the answers obtained after listening to Xg and Xd targets showed that a falling pitch remained a very strong option in favor of a low tone perception when the preceding tone was rising (as in P1). In the case of Xg, the falling pitch was perceived as a low tone even if it bore very high values (21 Hz above the natural values). This contrasted with the hesitations found on the Xd target, in which the second syllable pitch contour was also rising but in which the f0 of the last syllable was constant and even lower in pitch than in the Xg case.

5. Acknowledgements

The author would like to thank Dr. Denny Moore (Museu Goeldi/MCT, Brazil) for having revised this paper before submission, and who contributed to the choice of the tested stimuli. Dr. Moore was also a key person to contribute to build the project with the Suruí. Moreover, this research would not have been possible without the strong support of the Associação Metareilá do Povo Indígena Suruí and of its president Almir Suruí. The author would like to thank also the Suruí speakers and listeners who participated to the experiment as well as the institutional bodies which supported this research: either financially like the Endangered Language Documentation Program, SOAS (project IIP0136) or the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Pós Doutorado Junior Fellowship n°159747/2010-4), or for obtaining the necessary permits of research in Brazil (CNPq International, process No EX032/08-CR) in the protected Indigenous Territory (National Indian Foundation FUNAI). Finally, the members of the Linguistic Division of the Museu Goeldi (MPEG/MCT) are also thanked for their constant support in this research.

6. References