



HAL
open science

A study of the production and perception of ' in Tsimane'

William Havard, Camila Scaff, Loann Peurey, Alejandrina Cristia

► **To cite this version:**

William Havard, Camila Scaff, Loann Peurey, Alejandrina Cristia. A study of the production and perception of ' in Tsimane'. Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL), Nov 2022, Marseille, France. pp.1-8. hal-03852211

HAL Id: hal-03852211

<https://hal.science/hal-03852211>

Submitted on 14 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Production et Perception de ⟨'⟩ chez les Tsimane'

William N. Havard^{1, 2} Camila Scaff^{3, 1} Loann Peurey¹ Alejandrina Cristia¹

(1) Language Acquisition Across Cultures Team, Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'Etudes cognitives, ENS, EHESS, CNRS, PSL University, France

(2) Cognitive Machine Learning Team, INRIA, Paris, France

(3) Humam Ecology Group, Institute of Evolutionary Medicine, University of Zurich, Switzerland

william.havard@gmail.com, camila.scaff@cri-paris.org,

loannpeurey@gmail.com, alecristia@gmail.com

RÉSUMÉ

Le tsimane' est une langue parlée en Bolivie par plusieurs milliers de personnes et pourtant la phonologie du tsimane' n'a pas été décrite en détail. Avec ce projet, nous souhaitons faire un pas vers une meilleure description en nous concentrant sur un aspect de la langue qui nous semble particulièrement inhabituel : le son représenté orthographiquement par ⟨'⟩, traditionnellement analysé comme un coup de glotte /ʔ/. Nous avons émis l'hypothèse que, en termes techniques, ⟨'⟩ est réalisé comme un son glottal battu. Nous avons enregistré deux locuteurs adultes de Tsimane' produisant des paires (quasi) minimales impliquant ce son. Dans cet article, nous présentons des analyses centrées sur une syllabe extraite de six paires minimales : /ki-kiʔ/. Les analyses des spectrogrammes ont suggéré qu'un locuteur utilisait systématiquement la glottalisation de la voyelle et, dans une moindre mesure, une occlusion totale, alors qu'elles étaient ambiguës chez notre autre informateur. Cependant, la présentation des syllabes cibles à ces deux informateurs ainsi qu'à deux autres locuteurs adultes de Tsimane' a révélé des preuves claires qu'ils pouvaient clairement retrouver la syllabe voulue. Ensemble, ces données suffisent à écarter notre hypothèse initiale d'un flap glottal, puisqu'une fermeture n'était jamais évidente chez l'un des locuteurs, et suggèrent plutôt qu'un ensemble plus complexe d'indices acoustiques peut être à la disposition des auditeurs.

ABSTRACT

A study of the production and perception of ⟨'⟩ in Tsimane'.

Tsimane' is a language spoken in Bolivia by several thousand people and yet the phonology of Tsimane' has not been described in detail. With this project, we want to take a step towards better description by focusing on an aspect of language that we find particularly unusual : the sound represented in spelling with ⟨'⟩, usually analyzed as a glottal stop /ʔ/. We hypothesized that ⟨'⟩ is a glottal flap. We recorded two adult speakers of Tsimane' producing (near-)minimal pairs involving this sound. In this paper, we present analyses focused on a syllable extracted from six minimal pairs : /ki-kiʔ/. Analyses of the spectrograms suggested one speaker consistently used vowel glottalization and to a lesser extent closure, whereas these were ambiguous in our other informant. However, presentation of the key syllables to these two informants and two other adult Tsimane' listeners revealed clear evidence that they could clearly recover the intended syllable. Together, these data suffice to rule out our initial hypothesis of a glottal flap, since a closure was never obvious in one of the speakers, and suggests instead a more complex set of acoustic cues may be at listeners' disposal.

MOTS-CLÉS : phonologie, perception, production, adaptation des expériences de laboratoire.

KEYWORDS: phonology, perception, production, adapted lab experiments.

1 Introduction

There are few linguistic descriptions of Tsimane' (Sakel, 2011; Gill, 1999; Gill & Gill, 1999; Ritchie, 2017a,b), and none covers the phonology of Tsimane' very well. In this paper, we describe a small-scale research project aimed at studying the sound marked in the orthography as ⟨'⟩.

⟨'⟩ has been described to us as a glottal stop (Ritchie, undocumented personal communication, 2018), which has been well documented across languages. However, both distributional and phonetic evidence do not point to this really being a glottal stop. In distributional terms, glottal stops are hard to hear, so in most languages, they occur only at word and syllable onsets. When a language allows it in syllable offset, it also allows it in syllable onset – again suggesting that glottal stops will only appear in salient positions. In contrast, in Tsimane', ⟨'⟩ occurs *only* in syllable offsets. Moreover, it doesn't seem to be a "real consonant" distributionally, because Tsimane' allows very few consonants in syllable offset, and never allows two consonants together in that position. And yet, ⟨'⟩ can follow the nasal consonants /m, n/ or approximants /r, v/ at the end of syllables, suggesting that it doesn't "count" as a consonant from a syllable structure point of view.

Regarding phonetic evidence, based on preliminary discussions with Tsimane' research assistants and others, ⟨'⟩ does not sound like full glottalization of the preceding vowel or nasal (i.e., it does not sound like the vowel or nasal has a different vocal quality caused by the glottis vibrating in a specific way during the sound) but instead, that it is a "glottal flap/tap" (where the air is stopped at the glottis more briefly than for a stop, or perhaps not at all).

Based on these descriptions and considerations, we designed a small-scale study to be carried out during a 6-week visit to San Borja, a city close to traditional Tsimane' territory. Our original hypothesis was that ⟨'⟩ is phonologically a feature of the preceding sound (i.e., the vowel or nasal sound /m, n, r, v/ that immediately precedes ⟨'⟩) and phonetically a glottal flap/cued primarily by a change in the vowel or nasal formants. This hypothesis needed to be tested via acoustic analyses of speech samples and perceptual reports by informants.

Although there exist some speech corpora in Tsimane', none of them have the high audio quality required for our analyses nor for eliciting judgments by native Tsimane' informants. Therefore, our first aim during our short visit was to collect good quality audio recordings of a small number of informants producing the target sound ⟨'⟩ in a given context, as well as the same context without that sound. The second aim was to present extracts from those recordings, potentially edited to neutralize certain acoustic cues, to the same or other native Tsimane' informants, in order to assess the extent to which they rely on that acoustic information in their perception of ⟨'⟩ being present or not. Given time constraints, our perceptual studies do not include edited stimuli, but only natural stimuli. We return to this below.

2 Production study

2.1 Methods

2.1.1 Equipment and procedure

For the recordings, informants were seated in the same room as the experimenter and were provided with Tsimane' phrases to read, each containing one of the target words from our set of minimal pairs. We intended to use LIG-Aikuma (Gauthier *et al.*, 2016) to collect this data, using a simple smartphone. However, due to low audio quality (16kHz) as well as random recording failures, we implemented a computer-based version that re-implements the text-elicitation and respeaking modes and allows to record at a 44.1kHz sampling-rate.¹

Participants recorded sentences in a quiet room using a Dell Precision 3561 Computer running Ubuntu 20.04.5 LTS. Participants use JBL Quantum 300 headphones, equipped with a foam windscreens to record the stimuli. Headsets were connected to the computer via a USB audio cable adapter.

2.1.2 Stimuli

The words were chosen to elucidate the phonetic properties of ⟨ʔ⟩ in a minimally varying context, so we intended to find minimal pairs. Our analyses of the Tsimane' dictionary initially revealed a certain number of meaningful minimal pairs. However, in collaboration with our two informants, many of these items were found to have been mistranscribed in the dictionary, or to be variants of each other. Thus, in collaboration with our two informants, we identified 14 minimal pairs, and 6 near-minimal pairs (in which one more sound differed between the paired items). Additional work allowed us to identify more (near-)minimal pairs, and to construct meaningless minimal pairs. However, analyses here will focus on the 6 meaningful minimal pairs first identified where ⟨ʔ⟩ occurred in the syllable "qui" /ki/, so as to control for context and lexical status.

Each of the two items of each pair was recorded in five contexts : isolation, a natural phrase (drawn from the dictionary or grammar, or elicited from our informants) and three carrier sentences, where the position of the target word varies are used, so as to mitigate the effects of prosodic variations (e.g. "TARGET-WORD mo' nash peyacye' yu yi", which means "TARGET-WORD is the word I am saying"; "yu ra' yi TARGET-WORD jeñej peyacye'", which means "I will say TARGET-WORD as a word"; "yu ra' yi mo' peyacye' TARGET-WORD ", which means "I will say the word TARGET-WORD ").

2.1.3 Informants

This work was possible thanks to an extensive collaboration with two informants, both of whom are native Tsimane' speakers. They also use Spanish frequently, and this was the language in which all studies were conducted. They are both literate, and very experienced in collaborating with researchers. One of them had previously worked as an informant with a linguist; the other has over 10 years of experience as a translator and research assistant with a team of biological anthropologists. They grew up in two different villages, and our interactions with them in the transcription and analysis of Tsimane' materials strongly suggested to us that there are idiosyncratic or dialectal differences in

1. Code available at <https://github.com/William-N-Havard/williaikuma>

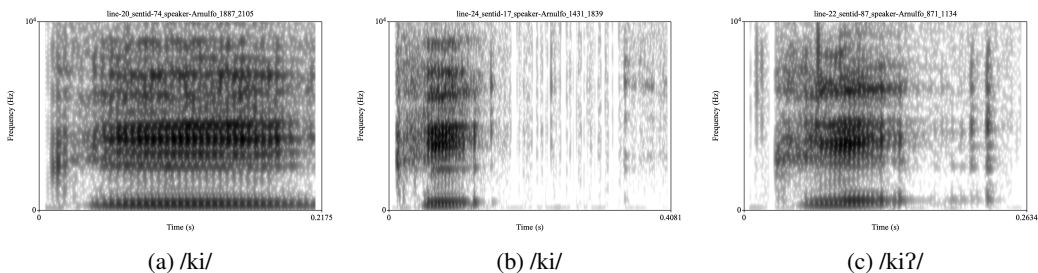


FIGURE 1 – A salient contrast between /ki/ (left) and /ki?/ (right), as well as an ambiguous case (/ki/ intended). The example on the left has no glottalization in the vowel or apparent glottal closure/release. The example on the right has both clear glottalization in the vowel and closure followed by release. The example in the middle seems to have some glottalization towards the end of the vowel, and what could seem like a very long closure and a release.

their production and potentially their perception. As we explain below, this was confirmed for the contrast of interest here.

This study was approved by CER U-Paris 2022-84-CRISTIA. Informants were compensated for their participation in this work, as well as in other ongoing work, at a rate of 145 Bolivianos per day.

2.1.4 Analyses

We call a "clip" each audio recording of one item in one context (isolation, natural sentence, or carrier phrase). First, we used Praat to inspect spectrograms of over 100 items (including ⟨ʔ⟩ or not, in a variety of vocalic and sonorant contexts) by our first informant, which made it obvious that checking for formant transitions visually would be very challenging. Instead, we focused on two aspects : (a) the presence of glottalization in the vowel, as defined perceptually as well as in the spectrogram (compare Figures 1a and 1c); and (b) the presence of a clear closure followed by a release (see Figure 1c). Third, we segmented 1. the syllable, starting before the release of the /k/ and ending with the vowel /i/ or, if present, a potential closure followed by release; 2. the vowel /i/ (from the first to the last voiced period); 3. if present, the closure followed by release if present. A Praat script was used to extract the duration of the vowel and, if an unambiguous closure was present, its duration.

2.2 Results

We annotated 136 clips from our two informants (68 clips each). Table 1 shows the prevalence of vowels judged to be glottalized, and closures judged to be present, for each speaker separately. From this analysis, it appears that one of our two informants consistently uses vowel glottalization (unambiguously present in 83% of /ki?/ clips, and unambiguously absent in 84% of /ki/ clips); and he often also signals a closure, although closures were ambiguous present in half of /ki/ clips, meaning that the speaker paused after /ki/. We found vowel glottalization status to be ambiguous in nearly half of the clips corresponding to our second informant. However, when glottalization was not ambiguous, it was clearly absent for /ki/, and more commonly present than absent for /ki?/. This second informant never produced unambiguous closures, regardless of the target.

| Target | Glottalization | | | Closure | | |
|--------|----------------|---------|-----------|---------|---------|-----------|
| | Present | Absent | Ambiguous | Present | Absent | Ambiguous |
| /ki/ | 1 / 0 | 31 / 22 | 6 / 16 | 0 / 0 | 20 / 23 | 18 / 15 |
| /ki?/ | 25 / 13 | 0 / 3 | 5 / 14 | 21 / 0 | 2 / 13 | 7 / 17 |

TABLE 1 – Perceptually-determined presence of vowel glottalization and closure (followed by release), as a function of target. Each cell indicates the prevalence separately for each of our two informants.

We then analyzed vowel duration separately for our two informants. In the first informant, an unpaired Welch two-sample t-test revealed a significant difference in vowel duration across /ki/ and /ki?/ : $t(64.3) = 4.3$, $p < .001$; 119 ms in /ki/ versus 80ms in /ki?/. In the second informant, the difference in vowel duration was not significant : $t(63.4) = 1.0$, $p = .32$; 171 ms in /ki/ versus 163ms in /ki?/. Finally, although we have nothing to compare this to, we thought it may still be informative to report on the duration of the closure (plus release) in our first speaker (since the second did not produce unambiguous closures) : The mean closure (and release) duration was 179 ms (range 64-298ms).

2.3 Discussion

Due to logistical constraints, we could only collect production data from two adult Tsimane' speakers, who frequently use Spanish. Nonetheless, they are both native and fluent Tsimane' speakers, highly experienced in research, and detailed-oriented, leading us to be confident that their production reflects clear speech. In these data, it appears to be the case that the contrast is stronger in one of our speakers than the other. Whether this is due to idiosyncratic or dialectal variation needs to be established in further work by interviewing more speakers. In the speaker where the contrast was clearly present, it was obvious in vowel glottalization, but also in the frequent presence of a closure.

3 Perception study

Given the natural variation in cue presence across our two speakers, we believe our perception study, based only on natural syllables (without artificial cue manipulation) can be quite informative. So as not to confuse listeners, we separated tokens by speaker, and collected perceptual reports through two separate listening studies.

3.1 Methods

3.1.1 Equipment and procedure

Praat was used to present the stimuli and collect judgments. Participants were tested one at a time in a quiet room. Participants sat in front of the computer and wore JBL Quantum 300 headsets at full power. There were three experiments presented in succession. The first was simply to acquaint participants with the procedure : two maximally different syllables (/ko^h/) and /yi/) were extracted from the speech of our first informant. These two clips were presented a total of three times in a random order. On the screen, participants could see the syllables "co'" and "yi", as well as a button that

allowed them to listen to the same stimulus up to three times before making a decision. We explained the procedure and stayed with them while they did this. We then answered any questions that they could have, and provided them feedback on the procedure if needed. The next two experiments were self-paced. In one, they heard all the stimuli for our first informant, and they heard all the stimuli for the second informant.

3.1.2 Stimuli

We extracted all /ki/ and /ki?/ syllables from the phrases analyzed above, except for two syllables of our second speaker that contained a click. Due to an error, some syllables were extracted several times with slightly different segmentations, resulting in 138 stimuli to be presented, 72 from our first speaker and 66 from the second.

3.1.3 Participants

Our two informants participated, as well as two other native Tsimane' speakers, who worked as research assistants for another project. Three of them were male, one female. All are young to middle-aged adults (one born in the 70', one in the 80's, and the last two, which includes the female participant, were born in the 90's). All come from different communities. This study was approved by the CER U-Paris 2022-84-CRISTIA. Participants were compensated 145 Bolivianos as part of their daily workday. The perception study as a whole took about 15 minutes.

3.2 Results

Table 2 shows the percent of correct responses (i.e., listener reported hearing a /ki?/ when the target was /ki?/ and vice versa) separated by participant, speaker, and whether the speaker intended /ki/ or /ki?/. It is obvious that all for listeners were overwhelmingly capable of retrieving the intended syllable. This is remarkable given that there was wide variation in implementation, as described above.

| Participant | Speaker 1 | | Speaker 2 | |
|-------------|-----------|-------|-----------|-------|
| | /ki/ | /ki?/ | /ki/ | /ki?/ |
| 1 (m) | 85 | 100 | 95 | 93 |
| 2 (m) | 95 | 97 | 100 | 83 |
| 3 (m) | 82 | 72 | 86 | 72 |
| 4 (f) | 60 | 100 | 70 | 69 |

TABLE 2 – Percent of correct responses, separated by participant (rows), stimulus speaker and target (columns). Participants 1 and 2 correspond to our first and second informants of the Production study described above.

3.3 Discussion

Despite variation in the input, participants were well above chance, suggesting they probably used a mix of cues, some beyond vowel glottalization and closure. We know that in other contrasts, a multitude of acoustic cues can correspond to several alternate gestures. Here, listeners could be relying to differences in formant structure, formant transitions, and even pitch levels.

4 General discussion

Our hypothesis was that ⟨ʔ⟩ was actually glottalization in the vowel or sonorant. Our production experiment revealed this was not necessarily the case, since one of the speakers didn't unambiguously glottalize or use modal voicing for his vowels. Moreover, although the other speaker did, he also produced clear closures and releases in a majority of his items. These data also allowed us to rule out clearly the hypothesis whereby the standard pronunciation for the contrast relied on the presence of a glottal flap or stop, since one of our two speakers never systematically and unambiguously use full closures.

The perceptual study showed that both speakers were signaling the contrast, since all four listeners (two who had participated in the production study, and two who had not collaborated with us on this project) were well above chance level for both speakers and both items of the pair. In the future, analyses using random forests on a variety of acoustic cues could help us understand the acoustic indices listeners were employing for each one of our speakers. In addition, analyses could inspect whether relevant acoustic cues vary as a function of the position in which the target syllable occurred. We noticed very strong reduction at the end of phrases, particularly for our first informant, and furthermore in this position closures are likely less salient. Our stimuli allowed the word containing the target syllable to occur sentence-initially and sentence-medially before a sonorant (vowel or semi-vowel), which should allow us to further study potential effects on the following sonorant.

The present study has several limitations. We were only able to study production in two native speakers, who also have extensive experience with Spanish, and it is unclear to what extent their use of this second language may affect their pronunciation. That said, there is no particular reason to suppose that Spanish could affect it, since glottal stops and vowel glottalization do not occur in any salient way in the local variety of Spanish. Another limitation was the use of elicited phrases, and the focus on one specific syllable, /ki/, as context. Future studies can employ the corpus of production we have collected to generalize this procedure to other contexts. We also hope to utilize other extant corpora to study this contrast in a wider variety of materials. Our perceptual study is also affected by a relatively small sample size, although results are quite clear in that they suggest the contrast is easily recovered by listeners. Nonetheless, sample size will be a limitation when attempting to use these data to infer which acoustic cues are most useful to listeners. Our data also suggest there is some idiosyncratic and/or dialectal variation, for which greater sample sizes and purposeful sampling of different geographic areas would be ideal. Finally, taking into account the goal of establishing what is the phonological category underlying this contrast, instrumental studies may be necessary, so as to recover the precise gestures that are made, although transporting such equipment to field conditions may be challenging.

Remerciements

We thank our informers, Arnulfo Cary and Manuel Roca, without whom this study would have been impossible. Funding for this study comes from the J. S. McDonnell Foundation (Understanding Human Cognition Scholar Award); European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ExELang, Grant agreement No. 101001095).

Références

- GAUTHIER E., BLACHON D., BESACIER L., KOUARATA G.-N., ADDA-DECKER M., RIALLAND A., ADDA G. & BACHMAN G. (2016). LIG-AIKUMA : A mobile app to collect parallel speech for under-resourced language studies. In *Interspeech 2016 (short demo paper)*.
- GILL W. (1999). A pedagogical grammar of the Chimane (Tsimane') language. *San Borja, Bolivia : New Tribes Mission*.
- GILL W. & GILL R. (1999). Chimane-English Dictionary. *San Borja, Bolivia : New Tribes Mission*.
- RITCHIE S. (2017a). Agreement with the internal possessor in chimane* : A mediated locality approach. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, **41**(3), 660–716. DOI : <https://doi.org/10.1075/sl.41.3.05rit>.
- RITCHIE S. (2017b). Posesión y Relaciones Gramaticales en Chimane. *Lenguas Indígenas de Bolivia : Teoría y Práctica*, p.7.
- SAKEL J. (2011). *A grammar of Mosestén*, volume 33. Walter de Gruyter.