International Journal of American Linguistics
Title: Phonetic structure of Yoloxóchitl Mixtec consonants

Reviewer: Sonya Bird

Summary: this paper provides an overview of some of the phonetic details of the Yoloxóchitl Mixtec (YM) consonants, in particular 1) consonantal duration as a function of manner of articulation and prosodic structure, 2) closure duration and VOT of voiceless stops as a function of place of articulation and individual speaker, and 3) nasal (and oral) duration in nasal stops and pre-nasalized oral stops as a function of place of articulation and prosodic position. Based on 3), the authors also argue for a phonological analysis of pre-nasalized stops in which they are viewed allophones of nasal stops, predictably occurring preceding oral (as opposed to nasal) vowels.

General comments: As the authors point out, detailed phonetic descriptions of speech sounds are so important to conduct, especially in the case of little-studied Indigenous languages. These are essential for us to further our understanding of phonetic typology, and also potentially serve as valuable resources for language revitalization efforts. Thus, in terms of general topic, the work undertaken by the authors is laudable. However, in terms of the specific content and layout, the paper lacks clarity and cohesiveness; it falls short of being either a comprehensive description of the YM sound system (as in the IPA Illustrations published by JIPA) or a focused study of a single topic of interest (e.g. the phonetics and phonology of pre-nasalized stops). For this reason, my recommendation is to reject it at this time, and to encourage the authors to either put together a description of the sound system as a whole (including the content of the current manuscript), or focus the paper on a more in-depth study of a *single* topic of interest, for example a phonetic and phonological characterization of YM pre-nasalized stops. Either way, I have included specific comments below, to help the authors move forward.

Note: I may have misinterpreted the intended scope of the paper (possibly being misled by the title). If it was the case that the paper was meant to be a report specifically on YM *consonantal duration*, based on prosodic structure, manner of articulation, and place of articulation, then the paper could possibly be revised and resubmitted, with a more explicit statement about the scope of the paper and how all the different component studies fit together. The presentation of the pre-nasalized stops in particular would have to be adjusted to make sure that it formed a cohesive package with the rest of the durational descriptions. In addition, many of the specific comments listed below would have to be addressed. For example, the various results sections would have to be revised substantially, in particular with respect to clarifying the rationale and appropriateness of the statistical tests used to analyze the data.

Specific comments:
However the authors choose to move forward with this work, I hope that the detailed comments below will be helpful. As I said above, I believe that the overall research project is a very valuable one, and certainly many of the components of the current manuscript are worth pursuing further.

p. 3 flow: insert paragraph break before “While debates like these…”, and get rid of paragraph break before “Furthermore, it is by no means certain…”

p. 3 bottom of page: In my opinion, the most important reason to conduct detailed phonetic work on lesser studied languages is the last reason cited: to gain a better understanding of the range of variability that exists in sounds that are represented with a single IPA symbol. I would emphasize this point most, in comparison with the other points that you make.

p. 4 first paragraph: I am not following this example. Given that you are contrasting the [t] sound in Hindi vs. English, it would help clarify your argument if you provided the underlying phonemes you are assuming: are you assuming that [t] corresponds to /t/ in Hindi but to /d/ in English? What is the evidence that the voiceless unaspirated stop in English is an allophone of the voiced stop? Are you saying that the second consonant in <stop> is underlyingly /d/? I don’t think this is the standard analysis.

p. 4 second paragraph, first line: see my general note on this above: you say here that the scope of your paper is “different consonant types” in YM. Is this really the case? Or is it a study of consonantal duration specifically? The content of the paper, following the introduction, does not seem to match your idea about what the paper is about.

p. 5 last line: grammar – there is no verb in the phrase “the larger groupings into which YM belongs”

p. 6 second paragraph: it might be useful to provide a map of where YM is spoken, or at least give a more detailed description of the geographical area in which it is spoken.

p. 8 “Glottalization is considered to be a couplet-level autosegmental feature...” I would leave out any explanation of what a couplet is here, and simply refer the reader to the later section in which you introduce and explain it. It is sufficient to provide the distribution of glottal stop in terms of segmental context, as you do, with concrete examples but without reference to the couplet. This will avoid confusion at this point.

p. 9 first sentence “While the focus of the current study is not … of YM consonants” : add “and their distribution”, since it’s specifically their distribution that you are discussing here.

p. 9 “Fourth, like most Mixtecan languages...”: this seems out of place here, since the realization of voiceless unaspirated stops is a phonetic characteristic rather than a phonological/distributional one. Can you leave this out, or move it elsewhere?

p. 10: I am not following your explanation of what the couplet is: What is the size of a couplet? Is it defined in terms of morae or syllables? “Couplet” implies “two” – is a couplet always two morae, or two syllables, and if so, is it the same thing as a “binary foot”? You say that “many words in YM consist of a bimoraic couplet” – is this the only possible size for a couplet? Later in the same sentence, you say “monomorphemic words are maximally trimoraic” – do trimoraic words include multiple couplets, or a single couplet (with extrametrical material of some kind)? Perhaps it would be helpful here to define the couplet using more standard prosodic terminology.

p. 10 "… the language differs from those varieties Pike and others describe in that monomorphemic words are *maximally* trimoraic": What is the size of words in other languages? Are they bigger in YM, or smaller?

p. 10, on the maximal shape of words: for CVCVV, on p. 8 my understanding was that CVV sequences were obligatorily CV?V. Is this the case? If so, how do CVCVV and CVCV?V words differ? Are CVCVV actually CVCV:?

p. 11 Table 3: it would be useful to provide a concrete example of each type of word. Also, am I interpreting this right that bimoraic words are never derived? Also, how does your classification here fit with the notion of "couplet" that you introduce earlier?

Note: given the scope of your paper, as far as I can tell the couplet is not critical to the remainder of the paper. Given this, it seems to me that you could avoid confusion by simply omitting the discussion of couplets altogether. This would save you the time/effort to explain couplets more thoroughly.

p.11 second paragraph "As a result of this difference": what difference are you talking about, the difference between YM and other related languages? The connection isn't clear here between short and long vowels and the prosodic structure of words…

p. 12 section 2 there is something unnatural about the flow here. You've already been describing some of the prosodic properties of YM, so it's not clear why previously these descriptions were subsumed into a single introductory section and now all of a sudden stress gets its own section. Perhaps you could end section 1 with a summary/reminder of your rationale for laying out the paper in the way that you did.

p. 12 first paragraph: it is still not clear what a couplet is, and how it is defined. What is the relationship between couplets and words? Some concrete examples would help here. In particular, examples that include the **prosodic structure** you are assuming (either in the form of a tree or with nested brackets).

p. 12 first paragraph "In YM, stress falls on the final syllable of the couplet": Is this word-finally? Or does the right edge of the couplet not always correspond to the right edge of the word? Again, some examples that include the prosodic structure you are assuming would be very helpful here. Also, do you have a reference for this statement, or your own empirical evidence? It would be nice to say something here about what the basis is for this claim.

p. 12 first paragraph, last sentence: you say that the main acoustic realization of stress is increased duration, and that final vowels are lengthened. Are you sure that this is stress rather than simply word-final lengthening? Do you have any other evidence that stress is active in the language?

p. 12 second paragraph, first line: can you give an example illustrating what you mean by "couplet-medial position"? Would this be the underlined C in [CV<u>C</u>V]?

p. 12 second paragraph, "For instance, consonantal lengthening occurs in the post-tonic (ultimate) syllable": Is this couplet-medial, i.e. ['CVC:V] – the onset of the final syllable? Again, it would be very helpful here to include examples with prosodic structure for clarity. Incidentally, consonant lengthening occurs in Athabaskan languages as well (in verbs at least), and the standard analysis is that it marks the beginning of the stem syllable in verbs. Could lengthening be used to mark morphological boundaries in YM?

p. 13 I think the research question needs to be stated more clearly, and more clearly placed in the context of the goal/scope of the paper here. See my comments above – perhaps clarifying the text leading up to the research question will be enough to clarify the research question.

p. 13 word-initial position in monosyllabic words, e.g. /ka^{3}a^{2}/: I'm wondering about the syllabification here, given the two separate tones. What is the evidence that these are monosyllabic rather than CV.V? If they were disyllabic, would you expect to be CV. ʔV? And/or, do you have native speaker judgments on syllabicity here?

On a related note: for your bisyllabic words, can you mark stress for clarity?

p. 13 methodology: Would it have been possible to collect data with trimoraic/trisyllabic words as well? If consonant lengthening marks stress, and if stress falls on the final syllable in CVCVCV words, then it shouldn't be lengthened in medial unstressed position [CV**C**VCV], right? Measuring medial, non-final (unstressed) Cs would provide you with a good baseline for comparing medial, final (stressed) syllable Cs.

p. 14 measurements: note that prenasalized stops and affricates also have internal components to them, which could have been measured. Is there a reason you measured the internal components for stops but not for the other consonants?

p. 14 analysis: One of your independent variables is *position*. Given what you've said about stress in YM, isn't this variable confounded with *stress* (stressed vs. unstressed)? Again, if you had trisyllabic words, you could tease apart these two variables, by contrasting medial stressed with medial unstressed.

p. 15 First model "While there seems to be a slight tendency… there was too much variability in consonant duration by class…": there was also a lot of variability within each consonant class, right? (see large error bars), which also would make significant results less likely.

In your statistics, you don't report on the interaction between *position* and *class*, did you test for this (as is standard with an ANOVA)?

p. 15 Second model: Again, was there an interaction between *word size* and *class*? It is important to test for interactions, because if they exist then the main effects are not necessarily reliable, i.e. they may only be "real" effects in a subset of the data in a given condition. For example, if the effects of *word size* and *class* are both significant, and there is also an interaction, then it's possible that word size is only a significant factor for a subset of classes, like *approximants* say.

So you'd then have to test for the simple effect of *word size*, for each *class* (for example), to see where significance actually occurred.

p. 16 Figure 3: it looks to me like "initial monosyllable" consonants generally pattern with "medial disyllable" consonants. Again, this seems to me to indicate a stress effect rather than a position effect. Along the same lines, in the discussion below Figure 3, you mention that the word onsets in disyllables in YM are shorter than one would expect (compared to other languages), but this doesn't seem surprising to me given that they are presumably also unstressed.

p. 17 second paragraph "Those authors compared the degree of consonantal shortening…": of which consonants, i.e. in which syllables? This needs clarifying, to make it clear if/how the YM data are comparable with the English data.

p. 17 "While they found little evidence of polysyllabic shortening in left-headed words": meaning that stressed syllables don't tend to get shortened? Am I interpreting this right?

p. 18 Figure 4: can you make it clearer exactly with consonants/syllables are being plotted? Is the medial disyllable stressed in both languages? Again, it essential here to be clear about the prosodic structure of the forms being compared, and about what is stressed vs. unstressed. In right-headed English words, the "medial disyllable" C is stressed right (re.PORT), where as in left-headed English words, the "medial disyllable" C is unstressed (MA.son)? I'm getting confused in this section about what you are comparing, and therefore am not convinced by the details of your comparison. I agree with your ultimate conclusion that when you hold the prosodic (stress) position constant, your YM results are no longer surprising, but I think your conclusion will be more convincing if it is clearer that you are making appropriate comparisons.

p. 19 section 3: As I mentioned previously, it's not clear to me how you've chosen which particular topics to focus on, and the result is that the paper seems to lack cohesiveness. Can you provide clearer road maps to the reader about how all the topics that you are covering tie together?

p. 19 second paragraph: do velar stops ever lenite as far as approximants, as in Spanish? It might be worth citing the Spanish literature here.

p. 20 last paragraph above 3.1 "and a test for examining the variable lenition of velars": didn't you say that there was no lenition in your data?

p. 21: can you say where your 427 tokens come from, more specifically? You have 10 words by 8 speakers = 80 tokens. So where does 427 come from?

p. 21 data analysis: on p. 22 you mention a third factor *position* did not have a significant effect. Did you include this factor in your analysis as well? If so it should be mentioned here.

Also, you have *component* as an independent variable. This seems counter-intuitive to me, can you provide your rationale for including it as an independent variable rather than a second dependent variable?

p. 22 "… we are particularly interesting in how stop components (closure, VOT) vary in percent duration with the stop place of articulation". Yes, agreed, but then it doesn't make sense to treat *component* as an independent variable, does it? I'm not following your data analysis. Perhaps I'm misunderstanding what your *component* variable refers to? It seems to me that you need to test separately for significant effects of place of articulation on 1) closure and 2) VOT, no?

p. 22 "… a significant interaction between STOP POA and COMPONENT on the percent duration of the stop components". Again, I'm not following what this means. Also, can you provide the actual statistics for the post-hoc tests?

p. 23 post-hoc results "All possible pairings were significant" … "though no significant differences were found for…": These two statements are contradictory.

p. 23 below Figure 6 "a strong, largely reciprocal effect": what does "reciprocal" mean here? Note that this paragraph is a repetition of what you've already said on the previous page. In my experience, the standard way of presenting results of this kind is to first present the descriptives (in the form of figures/tables), and then present the statistics.

p. 24 Figure 7: these are really nice data to include, it's great to have a sense of how overall patterns match up with individual speaker patterns. The figure would be easier to interpret though if it provided the data in terms of percentages rather than raw durations, so that it would be easier to compare across speakers.

p. 26 Figure 8: Are these Cho & Ladefoged's figures, or your figures based on their data? If their figures, do you need copyright permission to reproduce them?

p. 27 Did you include the [x] tokens in calculating your average durations? Specify this.

p. 27 "…where the tendency for closure to be shorter in velar stops results in the failure to achieve dorso-velar contact": this seems like a chicken and egg question to me, it's not totally clear what the causal relationship is between short duration and articulatory under-shoot: is it possible (although maybe not likely) that articulatory undershoot triggers shorter duration?

p. 28 "… while the palatal nasal surfaces only in the onset position of word-final syllables, e.g. **C**VʔV#..." So are VʔV sequences considered a single syllable? How are they syllabified by native speakers? On p. 10 you provide the maximal size/shape of YM words, my assumption was that CVCVʔV was trisyllabic, but is it disyllabic? I'm not sure it matters, but it would be good to be consistent (and clear) about this. Perhaps you can provide syllable boundary information (CV.CVʔV) in this kind of example for clarity.

Note: another argument for final stress in YM is that palatal nasals occur only in the last (stressed) syllable. Cross-linguistically, languages use the widest range of sounds from their inventory in stressed syllables.

p. 29 You distinguish ALLOPHONIC from HYPERVOICING accounts of pre-nasalized stops, but your description of them suggests that they are both allophonic, no? "The second view holds that prenasalized stops are surface **allophones** of voiced stops which have undergone hypervoicing". Can you clarify what you mean by "allophonic"? Or perhaps simply avoid using term "allophones" in reference to hypervoicing.

Also, I'm confused about the hypervoicing argument: looking at the inventory, there are no plain voiced stops. Conceptually, how can you have prenasalized allophones of voiced stops, if there are no plain voiced stops in the language?

p. 29 "Since there is a clear contrast between oral and nasal vowels in Mixtec…" You haven't actually provided evidence for this, can you include here an illustrative minimal pair? (or take out "clear")

p. 29 bottom: if the Hypervoicing account is right, and given the rationale you provide, wouldn't we expect the frequency of distribution of prenasalized stops to be velar > dental > bilabial (with velars being the most common)? I guess you'd have to also consider the distribution of the stop consonants themselves: are velar stops less frequent than the other stops? (note: I'm not actually sure this question makes sense, since there don't appear to be plain voiced velar stops in the inventory at all…)

p. 30 "Thus, on anticipates that there will be negligible differences in the duration of nasal and oral closure across different POAs": Or the same differences as in other contexts? Nasals themselves may differ in duration by POA, right? (in cases where they are produced simply as nasals). Is there any literature on this?

p. 30 For hypervoicing "one anticipates that the nasal portion of the nasalized velar will be relatively longer than the oral portion". Or maybe the whole consonant will be shorter? I'm not sure the predictions are quite as straightforward as they appear (see also my previous comment). Making predictions about voiced oral stops can be done fairly safely I think, given what we know about aerodynamics and articulation. I'm not sure that we can extend these predictions though to prenasalized stops.

p. 31 "Note that all vowels following a nasal consonant are obligatorily nasalized in YM, as we observe in the stem forms in (e-h)". Do you mean specifically e. and g.? This sentence is somewhat confusing: the wording implies that the oral-nasal vowel contrast is neutralized following nasal consonants, which in turn implies that nothing happens to the nasal consonants themselves. This is not the case though, right? It's not the following vowel that is affected by the preceding consonant, it's the preceding consonant that's affected by the following vowel, right?

p. 32 third line from the bottom “If prenasalized stops are consonant clusters…” isn’t this a 3rd possibility (the other two being the allophonic argument vs. the hypervoicing argument)? This is the first mention of pre-nasalized stops being clusters.

In general, I’m not completely following your hypotheses in this section, or the rationale behind them.

p. 34 “at the top of section 4.2”: Typo, this should be “section 4”

p. 35 data analysis “The first model examined the influence of POA and word position on the duration of each of the different components”. Was this model applied only to dentals (and bilabials)? Velars are restricted in terms of their POA, right? I’m not a statistics expert so I could well be wrong here, but as far as I know, it’s it impossible to model (using an ANOVA anyway) a set of data if one of the levels of one factor (here: velar POA) is restricted in terms of another factor (here, initial position)? Can you clarify this?

p. 36 Figure 10: [^{m}b ~ m] seems more variable than [^{n}d ~ n] across word positions, could this difference in variability be related to the distribution/frequency facts involved? The least frequent is also the most variable?

p. 37 “in many cases there was no clear oral closure preceding the burst release”: there was also no burst release, right? Can you clarify the wording here?

p. 37, discussion of cases with missing oral closure: it would be worth citing the literature on lenition in Spanish here I think, e.g. Martínez-Celdrán & Regueira (2008), Martínez-Celdrán et al. (2004)

p. 38 Figure 12: For ease of comparison, it might be useful to use percentages rather than raw durations. I wonder also if you could plot initial and medial consonants on a single plot, again for visual ease.

p. 38 “The fact that prenasalized stops here are of equal duration as simple nasals… suggests that they are unary segments and not clusters in YM”. Ok, but (as mentioned in a previous comment) this wasn’t one of the hypotheses you laid out to begin with: what do your data say about the ALLOPHONIC vs. HYPERVOICING arguments?

p. 38-39 “Rather, it seems that the nature of the alternations shown in Table 4 is more convincing in this respect”. Convincing in terms of what? Which hypothesis are you trying to argue for here? This isn’t clear.

p. 39 “In the HYPERVOICING perspective, one predicts longer oral closure duration for more anterior stops…” What about nasal closure duration, isn’t that what you were interested in?

p. 39 last paragraph “What the data does show is that velic raising may either slightly precede or be coincidental with stop burst release”. Clarity: Do you mean because there is often no oral closure?

p. 40 first paragraph: You refer here to prenasalized stops in Austronesian languages, and compare them to YM prenasalized stops. Of the Austronesian languages, you say: "In each of these languages, there is a process of perseverative nasalization on vowels following nasal consonants" and then say that YM exhibits the same pattern. But in Austronesian languages the prenasalized stops are phonemic and the vowels (nasal vs. oral) are allophonic, whereas in YM the vowels are phonemic and the prenasalized stops are allophonic, am I understanding this right? If so, are the data really comparable? You end this paragraph with "Moreover, the phonetic observations agree with the phonological patterns involving regressive voicing assimilation". You've lost me here, what does regressive voicing assimilation have to do with prenasalized stops?

p. 40-41 Discussion of the velar pre-nasalized stop: Can you say something more about the phonological status of this sound? It cannot be an allophone of the velar nasal, since the velar nasal is not in the YM consonant inventory, right? So are you considering it a phoneme (albeit a marginal/infrequent one)?

p. 41 "but the degree to which consonantal shortening occurs is greater than predicted from previous research": Is this still true when stress is taken into consideration?

p. 42 "Polysyllabic shortening is stronger in right-dominant words in English…": I think you could simplify the presentation here by referring simply to unstressed vs. stressed syllables, rather than to higher level phenomena like right vs. left dominance.

p. 42 bottom: You suggest that velar consonant lenition might be an areal feature. Is it possible that it results from Spanish influence?

p. 43 "the nasality of the following vowel is a stronger cue to the phonological contrast than the relative duration of nasal and oral closure": this is confusing, it implies that the contrast is in the consonants, but the argument is that this is not the case, i.e. the contrast is in the vowels, right?

p. 44 5.3 "In our case, the two gave quite different patterns" and "un-lenited stops were even rarer in spontaneous speech". This is the first mention of spontaneous speech – wasn't your study based strictly on elicited speech?

p. 45 first paragraph: again, refer here to the Spanish literature.

p. 45 second paragraph: Can you say something more concrete here about how an aerodynamic investigation would further your understanding of the YM consonantal patterns?

p. 45 last paragraph: this seems out of the blue here, unless the tonal system of the language has some bearing on consonantal patterns?

References

Martínez-Celdrán (2004). Problems in the classification of approximants. JIPA 34(2): 201-210.
Martínez-Celdrán & Xosé Luís Regueira (2008) Spirant approximants in Galician. JIPA 38(1): 51-68.