Acoustic Evidence for High Vowel Devoicing in Lezgi

Ioana Chitoran[1] and Khalil Iskarous[2]

[1]Dartmouth College, USA   [2]Haskins Laboratories, USA
ioana.chitoran@dartmouth.edu   iskarous@haskins.yale.edu

Abstract

This study uses acoustic analysis to determine whether unstressed pretonic high vowels in Lezgi are deleted or devoiced. We argue that the vowel gesture is not deleted, but it is overlapped and consequently devoiced by the preceding [s] gesture. We use spectral analysis to test the increased gestural overlap hypothesis. Three results support this hypothesis and consequently the devoicing interpretation: lower average energy in [s] before [u], higher energy in [s] before [i] in contrast to [a], and higher energy in [s] before unstressed [i] in contrast to stressed [a].

1 Introduction

This paper is concerned with the articulatory mechanism of a process affecting unstressed high vowels in Lezgi, a NE Caucasian language of the Daghestanian family. We present evidence that the high vowels do not undergo syncope (i.e., deletion of a full vowel gesture), as reported in the literature, but rather gestural overlap and devoicing, which can be explained by changes in gestural timing and the reorganization of the surrounding consonantal gestures.

A significant number of languages have been reported to exhibit high vowel devoicing in unstressed pretonic position, in the vicinity of at least one voiceless obstruent (see [1] for a survey). Among these, some, like Lezgi [2,3,4] also have morphological alternations involving stressed voiced and unstressed devoiced high vowels. Descriptions of Lezgi normally report a process of high vowel "syncope" whereby pretonic high vowels in a word-initial syllable are lost after voiceless obstruents. The syncopated high vowels are impressionistically

reported to be maintained as secondary articulations on the preceding obstruents.

(1) | singular | plural | |
     | sik' | sʲk'-ár | 'fox' |
     | tʰupʰ | tʰʷp-ár | 'cannon' |
     | tʃʰykʰ | tʃʰʉkʷ-ér | 'flower' |

The goal of our study is to determine, based on acoustic data, whether the process present in Lezgi is vowel syncope (i.e., deletion) or devoicing. We infer from the acoustic analysis as much information as possible about the articulatory mechanism in Lezgi.

Different types of spectral analyses have previously been done to determine patterns of oral coarticulation between fricatives and a following devoiced vowel [5,6]. We take our analysis one step further, inferring from the acoustic analysis the presence or absence of a vowel gesture in the pretonic environment. We argue that in Lezgi the high vowel gesture is not deleted, but it is overlapped by the adjacent consonantal gestures. The same overlap explains its devoicing. We use acoustic data to test our gestural overlap hypothesis, laid out in more detail in the next section. The results of our study thus contribute to a better understanding of articulatory-acoustic relations.

2 Hypothesis

We hypothesize that the reported syncope in Lezgi is due to increased gestural overlap between C1 and V, as a result of stress shift away from the vowel. The absence of stress on the vowel shortens the vowel gesture and affects the relative timing of C1 and C2, bringing them closer together. If the vowel is sufficiently overlapped it may no longer be perceived as a full vowel. Non-high vowels also tend to shorten

in the pretonic syncope context [7]. This may lead to future reanalysis of C1C2 as an onset cluster in Lezgi. If C1 has a glottal abduction gesture associated with it, at increased overlap this gesture will extend over a constricted, short vowel gesture, devoicing it.

The basis for this hypothesis is the proposal made by Browman and Goldstein [8,9,10] that syllable positions are defined by specific modes of coordination between the gestures involved. An onset consonant gesture (CV) is hypothesized to be coupled in phase with the following vowel, resulting in gestures that begin synchronously. The model predicts that when the vowel gesture is longer (the case of a stressed non-high vowel) the synchronous coupling still allows a full CV sequence to be perceived even at higher degrees of overlap. But when the vowel gesture is shortened (the case of an unstressed high vowel) more of it can be hidden.

We hypothesize that spectral properties of [s] as C1 will show differences in coarticulation with the following vowel, if the vowel gesture has not been deleted entirely. This prediction is borne out for Japanese [5,6], based on perception studies and acoustic analysis using LPC spectra.

We want to determine to what extent the same holds true for Lezgi. If the unstressed vowel has devoiced, but is not deleted, and is highly overlapped with C1, then we should be able to measure the acoustic consequences of the coproduction of C1 and the unstressed vowel. If, on the other hand, the vowel is deleted, we would not expect to find consequences of coproduction.

3 Data

The data reported on come from five native speakers of Lezgi, recorded in the village of Yargun in Northern Azerbaijan. The recordings were done in a quiet room, using an Edirol R-09 digital recorder, at 16 bit, 44kHz sampling rate. A cardioid dynamic microphone was used (Audio-Technica ATM63HE). All speakers are bilingual in Lezgi and Azeri, but speak Lezgi on a daily basis. Because the older speakers do not read or write Lezgi, no written lists were used. The speakers were asked to construct sentences using the target words. Each sentence was repeated 5-7 times. The target words contained the

vowels [i, u, a] stressed and unstressed (the syncope environment), all in words where C1 is [s]. The following environments were compared:
- [s] + [i], stressed and unstressed (syncope context): [sík'] / [sik'-ár] 'fox' (sg/pl);
- [s] + [u], stressed and unstressed (syncope context): [súp^h] / [sup-ár] 'soup' (sg/pl), [sút^h] 'land measure'; - [s] + [a] stressed and unstressed: [sál] / [sal-ár] 'garden' (sg/pl), [sáf] / [saf-ár] 'sieve' (sg/pl).

Data analysis consisted of spectral analysis of two windows in the [s] preceding the stressed or unstressed vowel. Previous studies on vowel devoicing [5,6] have used linear predictive coding (LPC) to estimate spectra of fricatives. However, it is now well known in the speech signal processing literature [11] that LC misses important information such as zeros in fricative spectra, and may also mischaracterize the location of poles. We therefore adopted the reduced-variance method of multitaper analysis [12], previously used to analyze fricatives [13]. In this method, spectral estimates are formed from several copies of the signal, each windowed with a prolate spheroidal window, and the results are averaged together, reducing the variance of the estimation. We used the multitaper spectrum analysis algorithm in the Matlab Signal Processing Toolbox, with 8 windows used in the averaging. Two windows were extracted from each fricative, both lasting 40 ms. The first one is two-thirds into the fricative and the second is the last 40 ms of the fricative. The speech was preemphasized before the analysis.

4 Results

Our goal is to investigate whether the unstressed vowel deletes or devoices by measuring the acoustic consequences of C1-V coproduction, if any. We sampled the [s] noise two-thirds of the way into the [s] and at the very end of the [s]. The choice of frames was based on the hypothesis that if the unstressed vowel does not delete, it would most likely affect the fricative in its last third.

Figure 1 shows multitaper spectra at 2/3 into [s] before [a] and [i] (black) and [u] (gray) for two subjects. The stressed and unstressed data are pooled together. To quantify the data, we calculated the first four moments for the spectra and then studied the effect of the various vowels on the center of gravity,

i.e. the first moment, as in [6]. However, moment-based analysis of fricatives has been shown to be problematic [14], since moments are perhaps too broad a parametrization to capture differences between fricatives or between the effects of different vowels on the same fricative. Such effects are often localized to certain spectral regions or prominences. Moments proved to not be very useful in studying our data precisely because the effects of different vowels on the fricative spectrum are local in frequency. It can be seen from the two plots that the spectra for the rounded [u] and unrounded [a] and [i] are somewhat similar at the lowest and highest frequencies. But from about 4kHz to 9 kHz, one can see a difference in the energy for the unrounded vs. the rounded vowels, with the latter having less energy, presumably due to a longer front cavity.
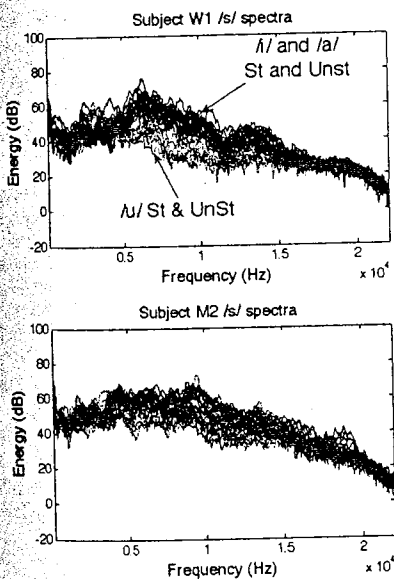
Moreover, since the data for [s] preceding stressed and unstressed vowels are pooled, it can be seen that [s] before the stressed and unstressed [u] pattern together against the unrounded vowels. To determine whether these patterns are significant and whether they extend to the other three subjects, the energy from 4 kHz to 9 kHz was averaged across frequency for each [s]. We also investigated the time course of the energy from 4 kHz to 9 kHz by sampling it at two-thirds into the fricative as well as at the last 33 ms frame of the fricative.

Figure 2 shows the results for all 5 subjects pooled together. For each vowel, means and standard deviations are shown across the subjects for [s] before the stressed and unstressed vowels at two-thirds (black) and end (gray) of the fricative. Energy is expressed in decibels.
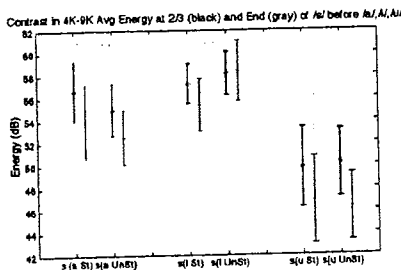

Figure 1: *Spectra at two-thirds into the fricative [s] before stressed and unstressed [a] and [i] (black) and stressed and unstressed [u] (gray)*

Figure 2: *Means and standard deviations shown with error bars for the average energy between 4 kHz and 9 kHz at two thirds (black) and end (gray) of fricative [s], before [a], [i], and [u] in pairs of stressed (left) and unstressed (right).*

It can be seen from Figure 2 that for the pooled data the energy is higher for [s] before [a] and [i] than before [u], for both the stressed and the unstressed cases. Moreover, unstressed [u] patterns with stressed [u]. We interpret this result to indicate coproduction between the rounded vowel [u], stressed or unstressed, and the [s]. If the unstressed vowel deletes there would be no labial gesture coproduced with the [s], that would lower the energy in this band. Under the devoicing hypothesis, however, the labial gesture remains and acts to lower the energy between 4 kHz and 9 kHz. Since the latter is in fact what

happens, we take this data to support the hypothesis that the unstressed vowel devoices but does not delete.

It can also be seen from Figure 2 that by the end of [s] there is a large difference between the 4 kHz-9 kHz energy in [a] vs. [i] for both the stressed and unstressed cases. Moreover, unstressed [i] seems to induce a *higher* average energy on the last frame of [s] than stressed [i]. Both of these results seem incompatible with the deletion hypothesis. We believe there are two possible interpretations of these results. One possibility is that lip spreading for [i] is coproduced with the last frame of [s], raising the average energy in the relevant frequency band. This would explain why [s] spectra before [i] would be raised in energy, but not why those for unstressed [i] are higher, on average, than those for stressed [i].

The second possibility follows directly from the overlap hypothesis. Due to increased overlap between C1 and the unstressed vowel, the tongue raising and fronting gesture for [i] is coproduced with the [s], fronting the constriction even further forward and encouraging excitation of higher frequencies.

The increased overlap hypothesis would explain all three effects: the lowering of average energy in [u], the raising of the energy in the [s] spectra right before [i] (in contrast to [a]), and the higher energy in the [s] before unstressed [i] (in contrast to stressed [i]).

10 Conclusions

In this study we examined acoustic data to determine whether unstressed pretonic high vowels in Lezgi are deleted or devoiced. We used spectral analysis of the [s] preceding the vowel to test our hypothesis, based on increased gestural overlap. Three results of [s] average energy converge to support an account based on gestural overlap (and thus devoicing).

References

[1] J.H. Greenberg. Some methods of dynamic comparison in linguistics. In J. Puhvel (ed.) *Substance and Structure in Language*. UC Berkeley Press, 1969.

[2] P.K. Uslar. *Etnografia kavkaza. Jazykoznanie. VI. Kjurinskij jazyk.* Tbilisi, 1896.

[3] S.V. Kodzasov. Fonetika. In A.E. Kibrik and S.V. Kodzasov (eds.) *Sopostovitel'noe izuchenie dagestanskix jazykov.* Moscow, 1990.

[4] M. Haspelmath. *A Grammar of Lezgian.* Mouton de Gruyter 1993.

[5] M. Beckman, A. Shoji. Spectral and perceptual evidence for CV Coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41:61-71.

[6] A. Tsuchida. Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence. *Working Papers of the Cornell Phonetics Laboratory* 9:183-222. 1994

[7] I. Chitoran, A. Babaliyeva. An acoustic description of high vowel syncope in Lezgian. *Proceedings of ICPhS* 16. 2153-2156. Saarbrücken, August, 2007.

[8] C. Browman, L. Goldstein. Gestural syllable position effects in American English. In F. Bell-Berti and L.J. Raphael (eds.) Producing Speech: Contemporary Issues. For Katherine Safford Harris. 19-34. AIP Press, 1995.

[9] C. Browman, L. Goldstein. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5: 25-34. 2000.

[10] L. Goldstein, D. Byrd, E. Saltzmann. The role of vocal tract gestural action units in understanding the evolution of phonology. In M.A. Arbib (ed.) *Action to Language Via the Mirror Neuron System.* 215-249. Cambridge University Press, 2006.

[11] D. O'Shaughnessy. *Speech Communication: Human and Machine,* Addison-Welsley 1987.

[12] D.J. Thomson. Spectrum Estimation and Harmonic Analysis. *Proceedings of the IEEE,* Volume 70: 1055-1096. 1982.

[13] O.S. Blacklock. *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods.* Ph.D. thesis, School of ECS, University of Southampton 2004.

[14] C.H. Shadle. Quantifying Spectral Characteristics of Fricatives. *ICSLP Proc.* 1996.