Articulatory and Acoustic Characteristics of Whistled Fricatives in Changana

Ryan K. Shosted
University of Illinois at Urbana-Champaign

1. Introduction

The whistled fricatives of Southern Bantu present a puzzle to both phoneticians and historical phonologists. The articulatory configuration of these sounds is the subject of some debate. Traditionally, linguists have claimed that labial protrusion, narrowing, or rounding imparts the whistled fricative’s characteristic whistle. The labial feature has played a role in determining the diachronic provenance of the Bantu whistled fricative (from proto-Bantu bilabial stops) (Maho 1999) and has gone on to inform related phonological analyses accounting for their synchronic behavior (Downing 2004).

Reports, which often conflict in their description of labialization, may indicate that labialization does not play a primary role in the production of the whistled fricative. Doke (1931: 47–48) refers to “general rounding of the lips” in the Shona (S10/S11–15) consonant <sv>. This is presumably in reference to both horizontal and vertical constriction, as well as protrusion. However, Maddieson (2003: 27) observes that for the whistled fricatives of Shona and Kalanga (S16/S16a), labialization “involves primarily a vertical narrowing of the lips with little or no protrusion.” Pongweni (1977) claims that the lips are closely rounded for Karanga’s (S14) whistled fricative. However, in Zezuru (S12), the primary articulator of the whistled fricative may be the “bringing forward of the lower jaw” while “there is no indication of lip rounding in the whistling sibilants” (Bladon et al. 1987: 40, 44).

Referring to the unique aerodynamic requirements of an edge-tone whistle (Chanaud 1970, Wilson et al. 1971, Coltman 1968), Shosted (2006) has argued that the proximity of the upper incisors to the tongue tip may play a role in producing the whistled fricative. He has further drawn comparisons to the non-labial, epiphenomenal whistle that has long been reported among speakers of English who wear dental prostheses (Kong & Hansen 2008, Cohen 2006, Runte et al. 2001, Petrović 1985, Sharry 1968, Silverman 1967, Rothman 1961). Shosted & Loucks (in prep.) are investigating the role of both linguopalatal configuration and lip kinematics in producing these widely-reported, though non-phonemic, whistles among speakers of American English.

In this paper, I present audio and video data collected from a single speaker of the Hlengwe dialect of Changana (S53/S511) which has two phonemic whistled fricatives. I suggest that labialization alone is insufficient for explaining the occurrence of the whistle. This highlights the ongoing need to study the lingual dynamics of whistled fricatives in the Shona (S10/S11–15) and

---

1 I am grateful foremost to the native speaker of Changana who consented to participate in this study and to Sharon Rose who arranged for me to meet him. This research has benefited from discussions with Torrey Loucks and audience members at ISSP 2006. It was supported in part by a Fulbright grant to the author and by institutional support from the Núcleo de Estudos de Línguas Indígenas Moçambicanas at the Universidade Eduardo Mondlane in Maputo, Mozambique. I am grateful to my Changana language consultants in Maputo, particularly Hélder Ngomacha, who assisted me with my original field work during my Fulbright year, 2000–2001. I am grateful also to Megan Osfar for her work on this project as a research assistant. I am solely responsible for any errors or omissions that remain.

1 I will refer to the referential classification systems of Guthrie (1967/71) and Maho (2003, 2009) using notation like (S10/S11–15) where the code to the left of the slash (e.g., S10) is used by Guthrie and the code to the right of the slash is used by Maho (e.g., S11–15). Where only one code is given, Guthrie and Maho are in agreement.

Tswa-Rhonga (S50) groups, as well as in two other non-Bantu languages, Tabassaran (Maddieson 2003) and Shehri/Jibbali (Johnstone 1984), where the sound reportedly occurs.

Figure 1. 10-ms interval extracted from the center of a whistled (top) and a plain alveolar fricative (bottom) uttered by a speaker of Changana. The periodic character of the whistled fricative is evident (the frequency is approximately 3.1 kHz). Both fricatives come from nonsense words where the consonant is flanked by the vowel [i].

Changana\(^2\) is spoken in five southern and central Mozambican provinces, as well as in neighboring South Africa and Zimbabwe (Sitoe 1996: vii). The language has a set of two alveolar whistled fricatives, voiced and voiceless. Representations of the sounds vary in the literature, including the use of a labialization diacritic and a vertical-pointing arrow situated beneath the main symbol to indicate whistling (ICPLA 1994, IPA 1979, IPA 1949). The sounds are represented orthographically as <sv> and <zv> in Shona (Hannan 1987) and Changana (Sitoe 1996), a convention that I will adopt here, including in phonetic transcription. Whistled fricatives in Shona have been described impressionistically in some detail (Doke 1931). Bladon et al. (1984) examined the acoustic and perceptual cues of the whistled fricatives in Zezuru (S12). Shosted (2006) has examined the acoustics of whistled fricatives in Tswa (S51). These authors generally agree that the whistled fricative is characterized spectrally by a relatively loud, narrow-bandwidth peak typically lower in frequency than the spectral peak of its non-whistled, alveolar counterpart. The whistle is easily observed in recordings of the sound, which has a periodic character and relatively simple harmonic structure (compared with that of a vowel or plain fricative) (Figure 1 above).

As mentioned previously, descriptions of the labial gesture associated with these fricatives vary greatly. In this study I hope to shed light on the production characteristics of Changana’s voiceless,

\(^2\) I follow Sitoe (1996) who uses the term ‘Changana’ in the primary lexicographical work on the language, and who refers to Hlengwe as a dialect of the language spoken in Inhambane Province (vii). It should be noted, however, that Changana (S53) is another term for ‘Tsonga’ which is sometimes used to refer to Changana, Tswa, and Ronga, three mutually-intelligible languages spoken in the city of Maputo and surrounding districts of Maputo Province (Paul 2009). Speakers of Tsowa and Ronga are apt to self-identify as speakers of Changana, though the converse is not necessarily true. In fact, Maho (2009) appears to classify Hlengwe as a dialect of Tswa, not Changana, counter to the intuitions of the native speaker who participated in this study and to Sitoe (1996: vii), who is the foremost authority on the subject. ‘Tsonga’ enjoys official language status in South Africa while the prefixed ‘Xichangana’ is formally designated as a national language of Mozambique (Lopes 1998). The name of the language is also listed as ‘Shangaan’, ‘Shangana’, ‘Shitsonga’, ‘Thonga’, and ‘Tonga’ (Paul 2009).
alveolar whistled fricative \([sv]\)^3 by comparing it to (coarticulatorily) rounded and unrounded productions of the plain alveolar fricative \([s]\).

2. Methods

A male speaker of Changana (Hlengwe dialect) from Inhambane, Mozambique, was recorded for this study. The speaker was about 25 years old and had lived in the United States for approximately five years at the time of the recording. Recordings were carried out in a sound-attenuated booth at the Center for Research in Language (CRL) on the campus of the University of California, San Diego. Like other Mozambicans from the country’s central region, the speaker is bilingual in Changana and another local language, Tonga (S62). Tonga has no whistled fricative.

For audio-video analysis, audio was recorded using the onboard microphone of a Canon FS200 digital video camera mounted on a tripod. The speaker was seated about two feet from the camera lens. Audio was recorded at a sampling rate of 48 kHz and bit depth of 32. Synchronized video was captured at 29 frames per second and a pixel depth of 720. The speaker held a compact mirror at the side of the lips, angled so that the degree of labial protrusion was visible in the image. Video was recorded as an MPEG file. Audio was stripped from the MPEG and exported in WAV format using Adobe Premiere Elements 4. For acoustic analysis, audio was recorded with an AKG C-520 head-mounted microphone attached to a Marantz PMD-660 digital recorder. Using this device, uncompressed audio was recorded at a sampling rate of 44.1 kHz and a bit depth of 16. The microphone was positioned approximately 10 cm from the corner of the subject’s mouth in order to direct the turbulent airstream away from the microphone.

The speaker repeated a list of nonsense VCV items where C varied between the voiceless whistled fricative \([sv]\) and the non-whistled voiceless fricative \([s]\). V1 and V2 were identical vowels: either \([i]\), \([u]\), or \([a]\), to test the coarticulatory effects of lip-rounding on the fricatives. During the video recording, the whistled tokens were repeated five times and the non-whistled items were repeated twice. During the audio recording the speaker repeated each item one additional time, so for acoustic analysis there were a total of six repetitions of whistled tokens and three repetitions of non-whistled tokens. A few natural-word tokens were also elicited and are included for illustration (Figures 2 and 3) but not in the boxplots, where more control over experimental variables is desirable (Figures 5, 6, 10, 11, and 12).

The audio signal was used to locate video images of whistled and non-whistled fricatives. Relevant frames were excised using Adobe Premiere Elements 4 and saved as JPEG files. The images were then rotated -11 degrees, cropped, and converted to black and white images. The brightness and contrast were adjusted using the auto-level and sharpen (= level 20) functions in Paint.NET 3.22. A vertical line was drawn from the superior vermilion border at the horizontal midline of the lips to the inferior vermilion border, also at the midline; a horizontal line was drawn connecting the corners of the lips. These lines were measured in pixels and then converted to cm using the length of the mirror (present in all frames) for calibration.

Fricatives were excised from the signals by hand, then downsampled to 10 kHz, and pre-emphasized at a factor of 0.98. Using the Signal Processing Toolbox in Matlab 2009b (7.9.0), multitaper spectral estimates of the fricatives were computed using orthogonal tapers specified from discrete prolate spheroidal sequences (based on Matlab’s DPSS function; Percival & Walden 1998, Blacklock & Shadle 2003, Blacklock 2004). This method was chosen to account for the stochastic quality of fricative noise, which gives rise to high variance in spectral shape from one time frame to the next. Least-squares regression lines were fitted to the frequency content of each fricative from 0.5 to 1.5 kHz \((S_a)\) and from 1.5 kHz to 10 kHz \((S_b)\) (Evers et al. 1998; cf. Jesus & Shadle 2002; see Figures 7–9). Each slope, measured in dB/Hz^2 (amplitudes have been normalized as dB/Hz, cf. Jones & Munhall 2003) represents the degree and direction of change in each region of the spectrum, e.g., a large positive value indicates that the energy in the spectrum is rising rapidly while a number relatively

---

3 I do not imply the production of a voiced labiodental fricative \([v]\) by using the \([sv]\) notation, even between brackets. The tiny vertical arrow used in the Extended IPA to indicate whistling is unfortunately difficult to typeset and to read.
close to zero indicates that there is little change in energy. Following Evers et al. (1998), the quantity \((S_a - S_b)\), or SlopeDiff, i.e., the difference in slope between the two regression lines, was calculated. Because \(S_b\) is typically negative, SlopeDiff is a large positive number when the spectrum is peaked at around 1.5 kHz; a relatively smaller positive number indicates the spectrum is relatively flat. Finally, when SlopeDiff is negative, this suggests that the greatest change in frequency lies above 1.5 kHz; in other words, the energy may be continuously rising across the spectrum (see Figure 7, left).

In addition, more traditional measures of fricative spectra, viz., center of gravity (CoG), standard deviation, and skewness were calculated for each multitaper spectral estimate (Forrest 1988). These measures are used to quantify the frequency at which the majority of spectral energy occurs (CoG), the distribution of this energy around the CoG (standard deviation), and the skewing of the distribution to the left or the right (skewness).

3. Results

![Frame 1](image1)
**Frame 1**
- I-S: 5.26 cm
- L-R: 7.22 cm

![Frame 2](image2)
**Frame 2**
- I-S: 4.90 cm
- L-R: 7.22 cm

![Frame 3](image3)
**Frame 3**
- I-S: 4.23 cm
- L-R: 7.16 cm

![Frame 4](image4)
**Frame 4**
- I-S: 3.65 cm
- L-R: 7.13 cm

![Frame 5](image5)
**Frame 5**
- I-S: 3.75 cm
- L-R: 7.02 cm

![Frame 6](image6)
**Frame 6**
- I-S: 3.75 cm
- L-R: 6.94 cm

**Figure 2.** Still frames of lips during the production of [esvi] in lè-svì ‘this-CL8’. An image was recorded about every 34 ms (29 samples per second). Simultaneous audio (Figure 3) indicates that the whistled fricative occurs most clearly in Frames 2–5. Inferior–superior (I-S) and left–right (L-R) distances for each frame are given.

Because multiple speakers were not available to participate, only descriptive statistics will be presented in this paper. Figures 2 and 3 present video and audio data for a single token of the word lè-svì ‘this-CL8’. Figure 3 is annotated in correspondence with the still frames presented in Figure 2. This enables us to observe the acoustic output associated with a particular labial configuration as it has been captured by the camera. The widths of the numbered ‘frame’ segments in Figure 3 highlight the discrepant sampling rates of the camera and the audio recording device: one video frame is captured approximately every 34 ms.

Figure 2 shows that lip spread decreased by only about 4% from Frame 1 (representing the final glottal pulses of the vowel [e]) to Frame 6 (the end of the high frequency peak in the whistled fricative [sv] plus the beginning of glottal vibration in the vowel [i]). Between these frames, distance between the upper and lower lips decreased by about 29%. Between Frames 5 and 6, where the amplitude of the whistle falls off dramatically, the decrease in left–right distance is only about 1% and there is no inferior–superior difference.
The spectrogram in Figure 3 shows a dark horizontal band (the whistle) appearing briefly in Frame 2, when the distance between the upper and lower lips is approximately 4.90 cm. The whistle appears again strongly in Frame 4, when the inferior–superior distance is 3.56 cm.

Figure 4 illustrates the labial postures assumed during the whistled and non-whistled fricatives in different vocalic environments. Non-whistled [s] can be strongly labialized, with a left–right distance of about 29% less than the unrounded [s] in [asa]. Similarly, the inferior–superior distance is 10% less for unrounded whistled and non-whistled [s]. In unrounded vocalic environments lip spread is about 9% less for a whistled fricative than for a plain fricative and about 17% less in terms of inferior–superior distance. However, for the [u] vocalic environment, [sv] has a greater left–right distance than [s]. For the fricatives in rounded environments, [usvu] and [usu], a different pattern emerges: the lips are slightly more spread for the whistled fricative (an increase of almost 9%). The inferior–superior difference is smaller for the rounded pair than for the unrounded pairs: the whistled fricative is only a bit more vertically constricted (by about 3%). The differences in VCV tokens across repetitions are given in Figures 5 and 6.
Figures 5–9 illustrate the spectral signatures of whistled and non-whistled fricatives in varying phonetic environments (flanked by unrounded vowels in Figures 7–8 and by rounded vowels in Figure 9). Each spectrum is accompanied by representations of the slopes ($S_a$ and $S_b$) described in Section 2; these appear curved because the frequency values have been plotted on a logarithmic scale which may better represent a listener’s perception of the sound.

For plain [s] in [asa] (Figure 7, left) the frequency peak occurs at about 9 kHz and $S_b$ is relatively steep. Because $S_a$ is relatively flat, this spectrum has a negative SlopeDiff ($S_a - S_b$, as defined in Section 2). The spectral peak for the whistled fricative (Figure 3) is much lower (approximately 1.4 kHz) and $S_b$ is less steep as a result (0.0359 dB/Hz²). This results in a large, positive SlopeDiff. Figure 7 also shows that the bandwidth of the spectral peak in the whistled fricative is much narrower than that of the plain fricative. Similar observations can be made for the plain and whistled fricatives illustrated in Figure 8, though the SlopeDiff of [s] in [isi] is positive.
Figure 7. Power spectrum of the plain fricative in the nonsense syllable [asa] (left) and the whistled fricative in the nonsense syllable [asva] (right). Gray lines (curved due to logarithmic x-axis) indicate the slope of spectral energy from (a) 0.5–1.5 kHz and (b) 1.5–10 kHz. SlopeDiff of [s] = -0.0018 dB/Hz²; SlopeDiff of [sv] = 0.0359 dB/Hz².

Figure 8. Power spectrum of the plain fricative in the nonsense syllable [isi] (left) and the whistled fricative in the nonsense syllable [isvi] (right). Gray lines (curved due to logarithmic x-axis) indicate the slope of spectral energy from (a) 0.5–1.5 kHz and (b) 1.5–10 kHz. SlopeDiff of [s] = 0.0163 dB/Hz²; SlopeDiff of [sv] = 0.0305 dB/Hz².

Figure 9 illustrates acoustic differences between a plain fricative [s] and a whistled fricative in a rounded environment, i.e., when flanked by the vowel [u]. The spectral properties observed in unrounded environments (Figures 7 and 8) seem to hold here, as well: a lower spectral peak for the whistled fricative, with a lower $S_b$. The narrow bandwidth peak can also be observed for the whistled fricative.
Figure 9. Power spectrum of the plain fricative in the nonsense syllable [usu] (left) and the whistled fricative in the nonsense syllable [usvu] (right). Gray lines (curved due to logarithmic x-axis) indicate the slope of spectral energy from (a) 0.5–1.5 kHz and (b) 1.5–10 kHz. SlopeDiff of [s] = 0.0111 dB/Hz²; SlopeDiff of [sv] = 0.0253 dB/Hz².

Boxplots of acoustic measures across repetitions of VCV tokens are given in Figures 10, 11, and 12. Differences between CoG (Figure 10) appear consistent across vocalic environments. CoG is consistently higher for the plain fricative than it is for the whistled fricative, indicating that most of the spectral energy in the whistled fricative is located in lower frequencies.

Non-zero skewness values indicate that spectral energy is not distributed equally about the mean. Positive skewness values indicate that energy is skewed to the left (i.e., there is relatively little spectral energy in the high frequencies). Negative skewness values indicate that energy is skewed to the right of the mean (i.e., there is relatively little spectral energy in the low frequencies). The skewness of [s] is consistently lower than the skewness of [sv] in the environment of [a] and [u], suggesting more high frequency energy is associated with [s]. For [isi] and [isvi], however, the two fricatives manifest roughly the same degree of skewing.

Finally, SlopeDiff (Figure 12) provides perhaps the most consistent acoustic differences between whistled and non-whistled fricatives. Low SlopeDiff values are indicative of high-frequency spectral peaks and/or a relatively flat spectral falloff (i.e., because $S_b$ is either positive or close to zero, the quantity $S_a - S_b$ will be a relatively small positive or even negative number). Conversely, a relatively peaked spectrum with a low-frequency peak is characterized by a high SlopeDiff value (i.e., because $S_b$ is negative, $S_a - (-S_b)$ will be positive). As Figure 12 suggests, whistled fricatives are characterized by relatively peaked spectra with low peaks, across vowel contexts.
4. Discussion and conclusion

The data presented here suggest acoustic differences between the whistled and non-whistled fricatives of Changana. These acoustic differences are similar to significant differences observed by Pongweni (1984) and Bladon et al. (1987) for Shona (S10/S11–15) and Zezuru (S12) and by Shosted (2006) for Tswana (S51): a lower spectral peak and a flatter high-frequency falloff (HiSlope) for the whistled fricative [sv] with respect to the non-whistled fricative [s].

The data regarding labial configuration present some new insights. A decrease in left–right distance is associated with the whistled fricative in unrounded vocalic environments. In the presence of [u] however, horizontal distance is minimized during the plain fricative, not the whistled fricative. Horizontally, the whistled fricative is less labialized than the plain rounded [s], even though the plain rounded [s] is not associated with the acoustic signature of a whistle. The decrease in inferior–superior distance is more orderly: in all vocalic environments the upper and lower lips are closer for [sv] than [s]. This is compatible with Maddieson’s (2003) observations of whistled fricatives in Shona and Kalanga (S16/S16a). However, for inferior–superior distance in the rounded environment the separation between [sv] and [s] is not as dramatic. To sum up, along the horizontal axis, [usu] is actually more labialized than [usvu]; along the vertical axis, [usu] and [usvu] may have comparable degrees of labialization.

The crucial question therefore deals with the acoustics of rounded fricatives, both plain and whistled. The two fricatives manifest lip-rounding but have very different spectral signatures. If labialization is the primary gesture responsible for producing a whistled fricative, why does the dramatically labialized [s] of [usu] fail to result in even a weak whistle? It seems uncontroversial to assert that alveolar fricatives manifest coarticulatory rounding in a variety of languages, including English. When labialization does occur systematically on sibilant fricatives, it generally lowers the spectral center of gravity, as with the alveopalatal fricative /ʃ/ in French and English (Abry et al. 1979,
If labialization is primarily responsible for the whistled fricative’s characteristic whistle, why are labialized sibilants not produced with an epiphenomenal whistle, e.g. in English and French? While a future study will hopefully consider the labial and lingual dynamics of Southern Bantu whistled fricatives in much greater detail and with much more data, for now it is only possible to hypothesize, based on the observations presented here, that Changana’s whistled fricative results from a unique linguopalatal rather than labial configuration.

However, this is not to deny the influence of labial configuration on the acoustic output. It may be the case that the acoustic whistle produced, e.g., by a retroflex tongue position, is enhanced by vertical narrowing of the oral aperture, for instance by lowering the spectral center of gravity as in English and French. However, it should be noted that the Zone S languages also have an alveopalatal fricative whose spectral peak is roughly comparable in frequency to that of the whistled fricative (Bladon et al. 1987: 48). The hypothesized retroflexion of whistled fricatives requires documentation. In Changana (S53), the whistled fricatives are transcribed as retroflexes (Sitoe 1996). Carter & Kahari (1979) also mention retroflexion for the whistled fricatives of Shona. Finally, Laver (1994: 252) has observed that “[a] retroflex fricative stricture at any place of articulation [may result in] a momentary whistle… though this will only happen at particular flow rates.” This suggests that further study of the aerodynamics and articulation of whistled fricatives is warranted. It seems possible to produce whistled sibilants with or without a retroflex lingual gesture. The key may be to create a relatively large posterior cavity in which vortices shed at the vena contracta have sufficient space to establish a feedback path, thus emphasizing the cavity’s natural resonance (Shadle 1983: 149).

Study of the whistled fricatives in Zone S, including their origins, may help linguists understand the relevance of secondary articulations in sound change. Ladefoged & Maddieson (1996: 354) define a secondary articulation as “a lesser degree of stricture accompanying a primary articulation of a higher degree” and suggest that “secondary articulations are always approximant-like in nature”. Ladefoged and Maddieson (1996: 368) observe that “[t]oday’s secondary articulations may be the primary articulations of the future” and refer explicitly to palatalization in Slavic as an example. However, Maddieson (2003: 19) remarks that the proposed development of Bantu whistled fricatives from historical bilabial consonants (Maho 1999) is “particularly unusual” because a historical process appears to have “reduce[d] the original [labial] consonant to a secondary feature”. So, while secondary articulations often become primary historically (e.g., through the phonologization of coarticulatory effects that lead to “blended articulations”), primary articulations are not as likely to become secondary, as appears to be the case in some Bantu Zone S languages. The phonetic mechanism underlying this interesting change merits further investigation.

References


I have observed epiphenomenal whistling in the retroflex fricatives of Q’anjob’al, a Mayan language of northwestern Guatemala.


edited by Eyamba G. Bokamba, Ryan K. Shosted, and Bezza Tesfaw Ayalew

Cascadilla Proceedings Project Somerville, MA 2011

Copyright information

© 2011 Cascadilla Proceedings Project, Somerville, MA. All rights reserved
ISBN 978-1-57473-446-1 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper. Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as: