Voice qualities and inverse filtering in Chong

Jerold A. EDMONDSON
University of Texas at Arlington

1. Introduction

Chong is a Mon-Khmer language of the Pearic subgroup spoken in Chantaburi Province, Thailand and in adjacent areas in Cambodia, its population is unknown. During August 1993 I had an opportunity to study the language at the local secondary school and health center of Khlongphlu Subdistrict, Khawkhitchagut District of Chantaburi Province for a period of about one half day as a part of a fieldtrip organized by the Institute of Language and Culture for Rural Development, Mahidol University at Salaya, Nakhon Pathom, Thailand.[1] On the trip we interviewed seven or eight male Chong speakers from ages 13 to 70.

Chong is a particularly interesting language for study as it has four "registers" and these registers involve four different voice quality contrasts.[2] In that sense Chong along with Bai, a Tibeto-Burman language of Yunnan Province, are to be regarded as languages with the richest use of these kinds of contrastive features known to date.

Chong has been reported on extensively by Huffman in "The Phonology of Chong" (1985) and in even finer detail by L-Thongkum (1992) in her paper "An Instrumental study of Chong registers". In this latter work L-Thongkum used a number of instruments and techniques to measure formant frequency, power spectra, fundamental frequency, duration, and intensity. See also Diffloth 1992.

My results basically confirm findings reported by L-Thongkum, but my method provides a different lens on the remarkable phonetic gestures used in the production of some Chong sounds. This study will employ the technique of inverse filtering to determine the dynamic modes of vibration of the glottal folds during speech. I will not discuss features of Chong other than certain aspects of the voice quality contrasts, as my time with the Chong in Chantaburi was not sufficient to carry out any broader phonological study.

---

[1] The author wishes to express his utmost gratitude to the Institute and to Acharn Suwilai Premsrirat and to Acharn Somsonge Burusphat of the Institute for arranging this visit. Thanks are also due Dr. Kenneth J. Gregerson for suggestions on an earlier version of this paper.

[2] By voice quality I mean the "overall auditory coloring of a given voice with both laryngeal and supralaryngeal factors contributing to it." (Laver 1980).

Huffman (1985:355-7) gives an overview of the kinds of Chong described, noting that Martin 1974 speaks of [Coŋ həəp] and [Coŋ Lɔɔ] and that the latter is quite like the kind which Huffman had studied at Ban Thung Saphan.[3]

2. Inverse filtering

2.1 Methods

According to the Source-Filter Theory of Speech Production (cf. Fant 1961, Fant and Sonesson 1962, Lindqvist 1965), the air pressure wave exiting the mouth is said to be the coproduct of *glottal source wave, resonance filtering* of a tube-like vocal tract, and *radiation* out into the atmosphere. By using the Rothenberg Mask and the tunable filters in the MSIF-2 Hardware Unit in the inverse filtering mode (cf. Figure 1), one can cancel the influence of the vocal tract configuration and recover the glottal source wave. At the same time, the mask allows us to capture all the airflow discharge into the atmosphere and, consequently, leaves no additional radiation effect to be considered. In order to remove these vocal tract components from the air pressure wave, the filters for the first and second formant frequencies (F_1 and F_2) are adjusted as the subject produces a modal voice syllable until the visually most completely filtered profile is obtained, i.e., one in which the resulting wave form resembles the idealized model depicted in Figure 1.[4] In this process one begins with the values of F_1 and F_2 obtained from a sound spectrogram, but precise tuning is still necessary as formant values are subject to considerable bandwidth and frequency variation and the slightest deviation from the resonance poles (formant frequency) leaves some of the resonance unremoved. A second problem is that even after the filters are set for the optimal F_1 and F_2, there is the added complication that the vowel formants in the test utterances might change over the course of the syllable or even that there was some interaction between the formants and the voice quality so that the settings might be correct for modal voice quality but they would be incorrect for others. Despite this potential for error, the results we obtained suggest that our simplifying assumptions—the language has pure or unchanging vowel and the voice quality-vowel quality interaction is minimal—were largely justified. Once the filters had been determined, we then maintained these settings for all other voice quality types so that comparison to the modal voice settings would be possible; in this sense modal voice was taken as the standard of comparison.

---

[3] Curiously, from the word for 'tongue' *kətáak* (as collected by Martin 1974), we would conclude we had studied Coŋ həəp. Despite the fact that our speakers produced *kətáak* instead of *táak* or *taaʔk*, we believe from the geographic location that we were studying a type quite similar to Huffman's.

[4] The equipment I used does not allow higher formant contributions to be removed from the signal. These may be responsible for Chong filtered waves (Figures 3-7) departing from the ideal (Figure 2) given below.

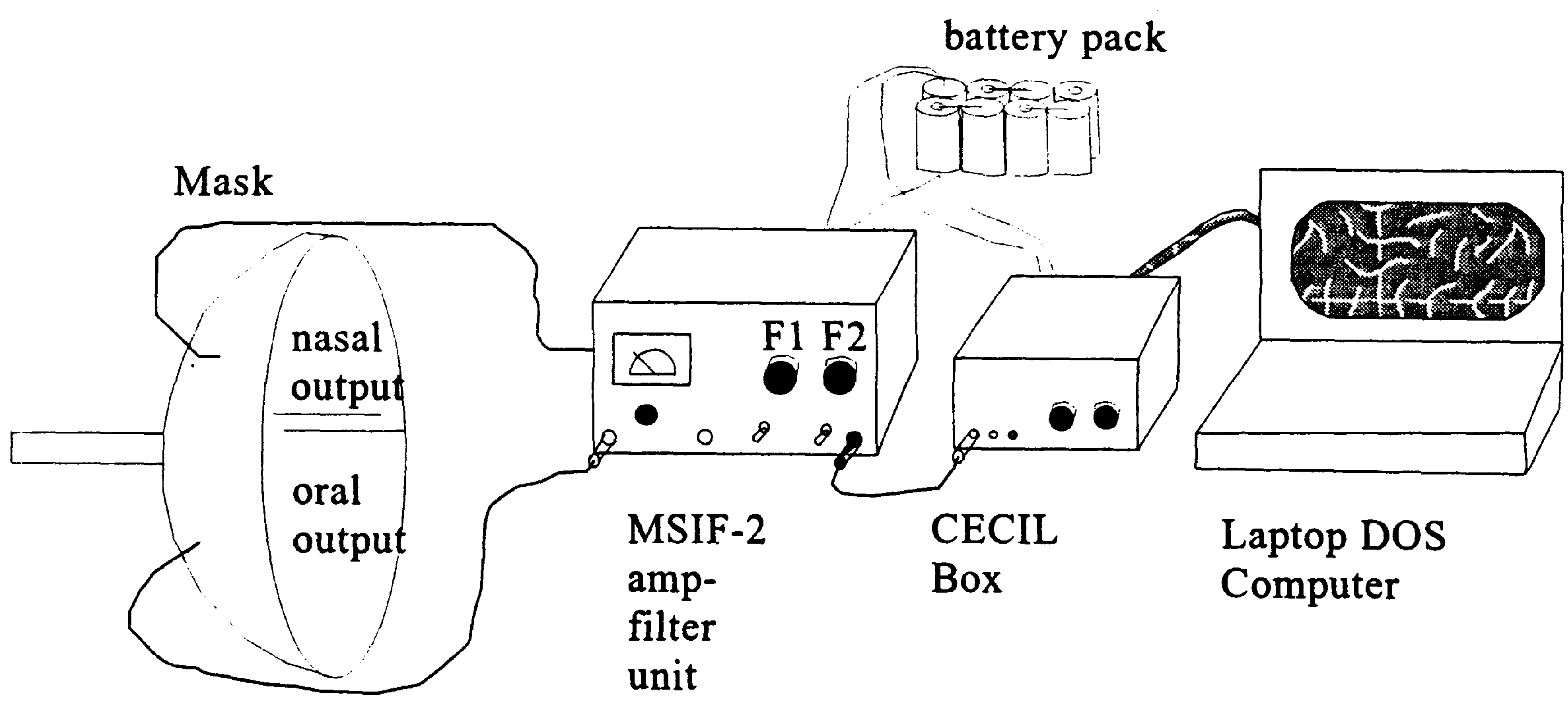

Figure 1. Inverse filtering using the Rothenburg Mask

The net result of the capturing mask and properly tuned Hardware Unit interfaced as in Figure 1 is to display the vibratory mode of the original air pressure wave on the computer screen. The mask allows us to measure both the steady-state eggressive airflow exiting the mouth (called the DC component) as well as the varying air pressure wave of glottal vibration (called the AC component). An ideal glottal air pressure wave for modal voice has the configuration illustrated in Figure 2 (Lindqvist 1965, Laver, 1980).

As can be seen in Figure 2, the glottal waveform has a dynamic opening phase, a dynamic closing phase, and a final static closed phase. It is noteworthy that opening phase in modal voice is relatively slower than the more rapid closing phase. It is also clear that the pulse height off the baseline is composed of two parts: (1) a varying AC part and (2) a steady-state DC part.

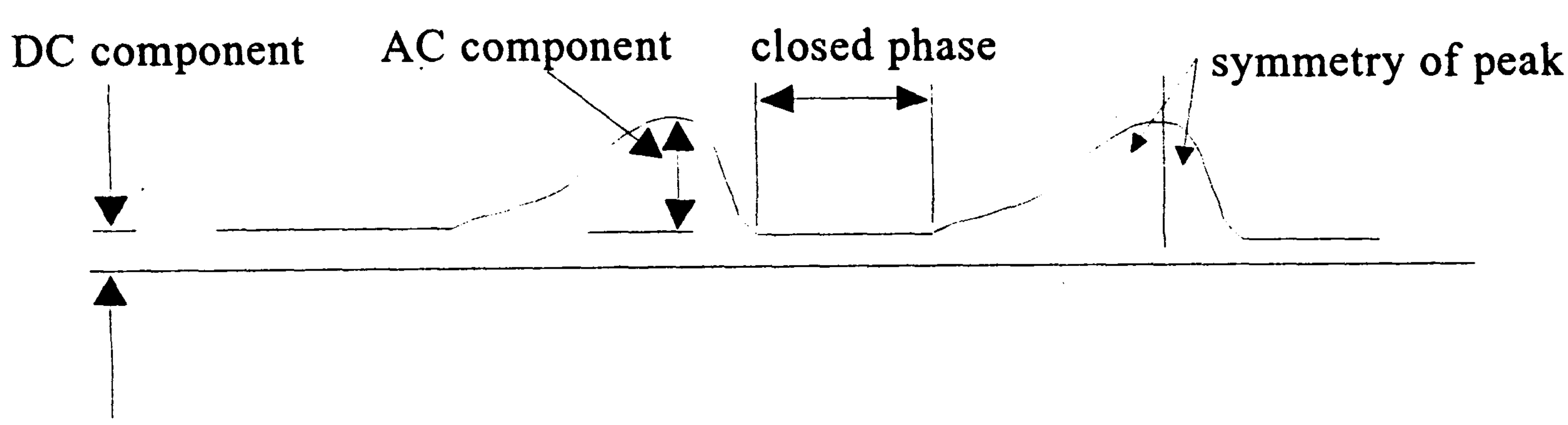

Figure 2. Idealized glottal waveform

3. The Four Chong registers

Like L-Thongkum, I found that not all speakers produce the phonological contrasts in exactly the same way. Our first subject was a thirteen-year old boy. He seemed mostly to use pitch and intensity to form the major contrast in his speech. A good example of such a contrast was *taŋ*˥˩ 'before' vs. *roŋ*˧˩ 'melon'. However, once I became familiar with the voice quality differences of many other subjects, I could easily perceive the commonalities between the boy and the older speakers.

Our second subject—who we will call Khun Bun—is the one I will report on here. The gravelly harsh nature of his speech in general was very apparent to us even when he was speaking Thai. In Chong, his utterances were spectacular. He possessed the very audible contrasts that were similar to but perhaps more extreme than those of others.

L-Thongkum describes the four salient contrasts as:

Register 1. Modal voice at syllable onset and throughout the syllable.
Register 2. Modal voice at onset followed by creaky voice.
Register 3. Breathy voice at onset.
Register 4. Breathy voice followed by creaky voice.

Our recordings of Khun Bun's speech did not seem to show *creaky voice*, as I understand that term to mean a voice quality with low fundamental frequency and lax passive and lax active tension with strong medial compression of the vocal folds (Laver 1980). In this speaker the F_O was raised over that found in modal voice. This voice quality was also tense and not lax. I would characterize the voice quality I heard as *raised larynx* with extremely *retracted tongue position* and *harsh voice quality*. I arrived at these designations by comparing several utterance tokens of the Chong speakers to the audio tape by Professor John Laver that accompany his 1980 book. I then judged which of the samples most closely matched the quality of Khun Bun's voice. I found Laver's example number 2 "raised larynx voice", if it were mixed with example number 36 "harsh voice", would be perceptually most like the utterances of Khun Bun. Features of the airflow recordings confirm this assessment.

There was a second unmistakable feature to Registers 2 and 4. Just as L-Thongkum has described, there was a strong change of glottal settings over time that I will call *glottal wringing*, a rapid dynamic constriction of the vocal folds, gradually approaching virtual closure of the glottis about two-thirds of the way into the syllable, only to be followed by a release of the constriction and slight opening, but not to the same degree of opening as at onset. L-Thongkum also noted a drop of fundamental frequency in R2 and R4 (cf. her Figure 7), suggesting that, "The insertion of laryngealisation or creakiness seems to be the cause of the falling F_O contour." (150).

I found *lowered larynx* and *lax or breathy voice* in the same environments as L-Thongkum (1992). In the citations below I will transcribe glottal wringing with a tilde under the vowel as in creaky voice. Breathy voice is indicated by two dots subscripted below the vowel.

Once we had listened to Khun Bun's speech, it was easy to perceive these features in the other speakers. In some cases they had raised larynx and glottal wringing but failed to have the harshness, but it was clear they were aiming for a similar auditory target.

First register (R1) syllables in Chong are illustrated (Figure 3) by the second syllable of the word 'duck' *kata*. Syllables with this modal voice quality have neither the elevated airflow nor the large-amplitude glottal vibration characteristic of breathy voice. At the same time they lack the pencil thin lines of a "wrung" syllable. The wave travels across the instrumental page at about the same level throughout. There is no dynamic change of either the AD (wave amplitude) or DC overall constant airflow. In a sense it serves as a standard of comparison for the other three voice qualities.

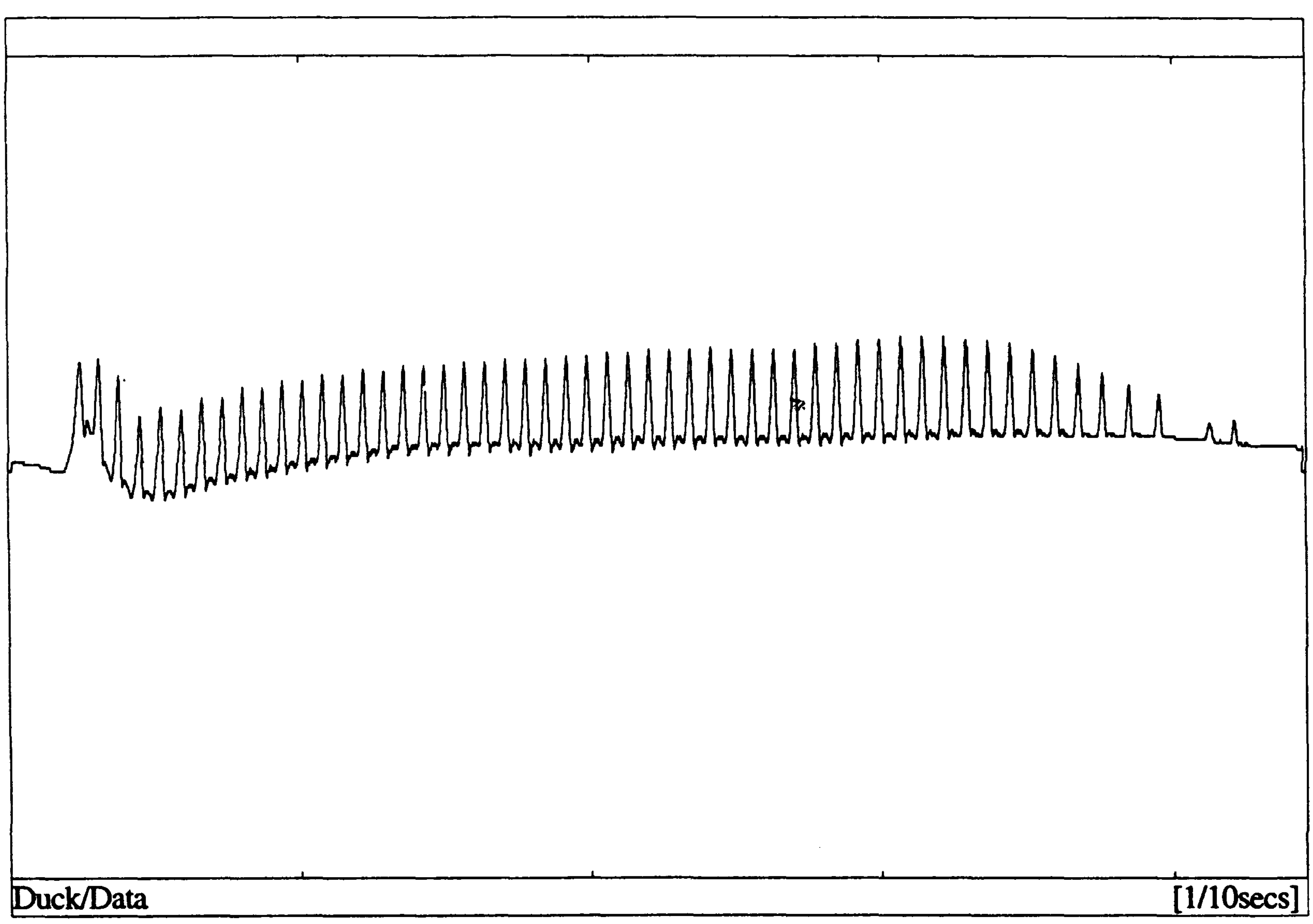

Figure 3. Modal voice in Chong illustrated by the glottogram for the second syllable of the word *kata* 'duck'

The second register (R2) has modal voice at onset followed by glottal wringing, as exemplified by the Chong word *kətaa̱k* 'tongue'. As can be seen in Figure 4, the airflow wave at syllable onset exhibits neither excessively large amplitude, nor does it show a large increase in general airflow. At a point about one-half to two-thirds through the syllable the AC component of the wave begins to change rapidly; the amplitude of the glottal vibrations becomes progressively shorter and shorter while the regularity of the glottal pulses disintegrates appreciably. This change transpires while the overall slope of the wave slightly rises, suggesting that the active tension is rapidly increasing, at the same time the passive tension is slowly decreasing. At the time of the greatest constriction some double peaked waves start to appear along with considerable irregularity in the waveform configuration indicating ventricularization or the involvement of the false vocal folds in the sound production (harshness). At the very end of the syllable there is again a growth in amplitude as the progress of squeezing reverses.

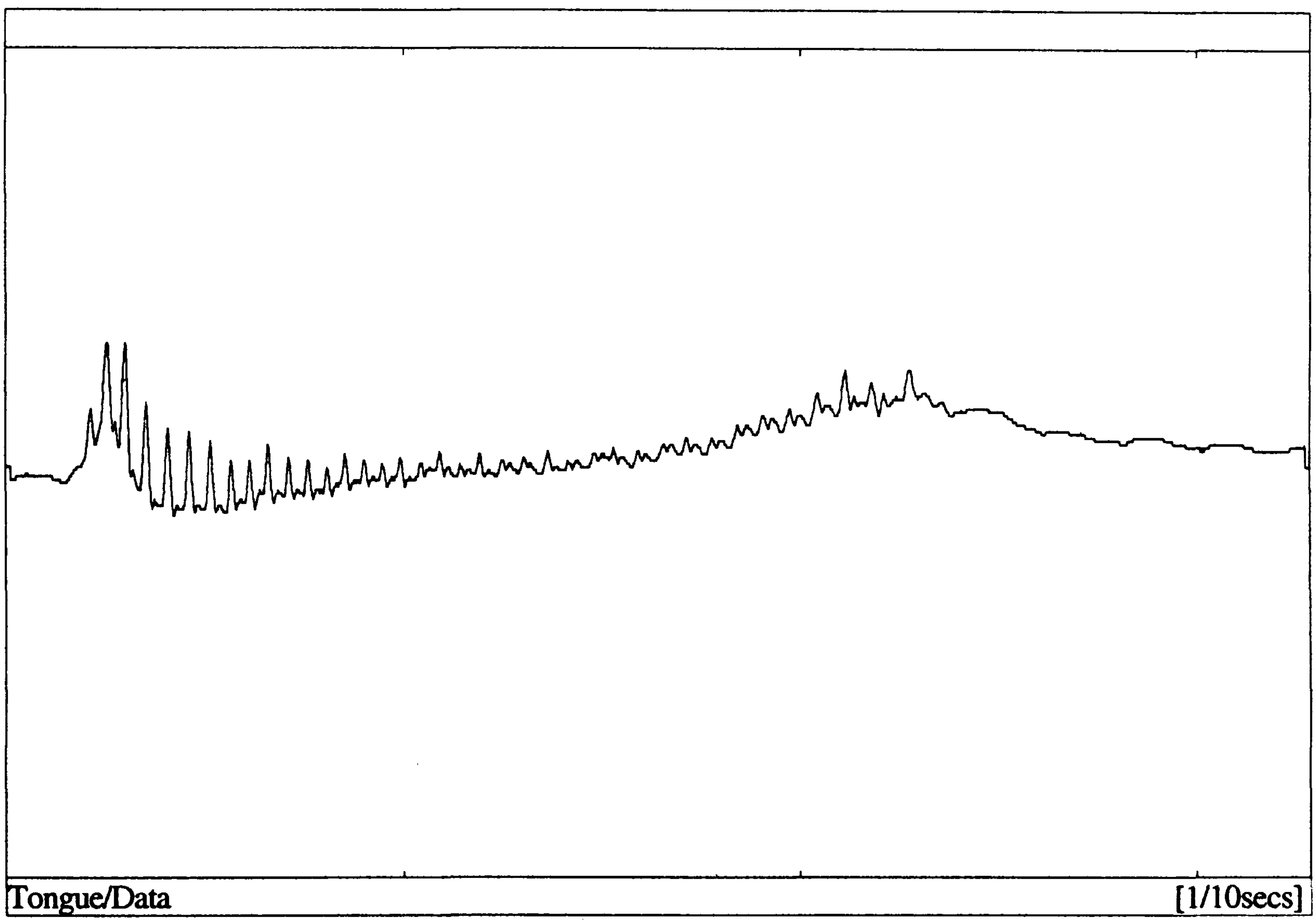

Figure 4. Chong *kətaa̱k* 'tongue'
Note: the severely throttled airflow during the midcourse of the syllable.

The third register (R3) has breathy voice onset, as found in the syllable *sa̤a̤* 'together'. In Figure 5 the increased flow of air at syllable onset is very evident. The glottal pulses demonstrate near symmetric vocal fold vibration at the beginning of the syllable with very little closed phase in the period. There is also a large increase in the overall airflow that peaks at the point vowel voicing commences. After a half dozen periods this profile changes to the modal voice configuration of well defined closed phase between pulses.

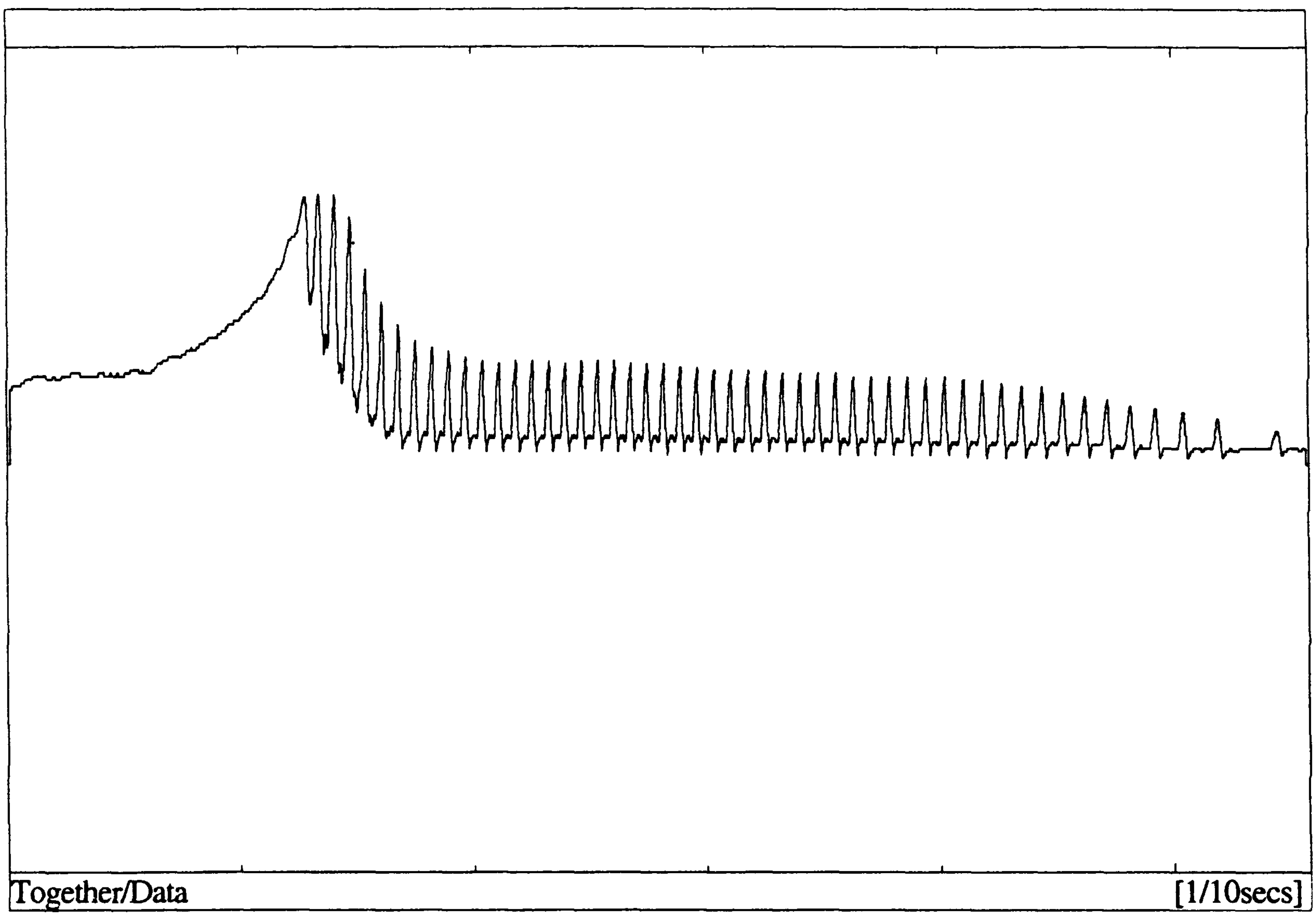

Figure 5. Breathy voice onset in the Chong word *sɑ̱ɑ̱* ‘together’

The following two figures show the glottograms for two examples of Register 4 contrast. The first of these is the word *mɯ̤ɯ̰t* ‘messy’ and *ɲɑ̱ɑ̰m.* ‘fingers’. In both these words there is strongly increasing DC airflow after onset of the voiced part of the syllable, as is seen in waveforms with general positive slope (travel “uphill”) to a point well off the baseline. Also the AC amplitude of vibration is quite large and very symmetric about the peak for the first 10 periods of the wave. As the syllable progresses this AC amplitude is diminished under the influence of dynamic throttling and irregularity in the wave shape increases. Even though the DC level is elevated above levels at onset, there is, nevertheless, near stoppage of the AC component about two-thirds of the way through. Finally, there is a detensing of the vocal folds, resulting in an increase of wave form amplitude.

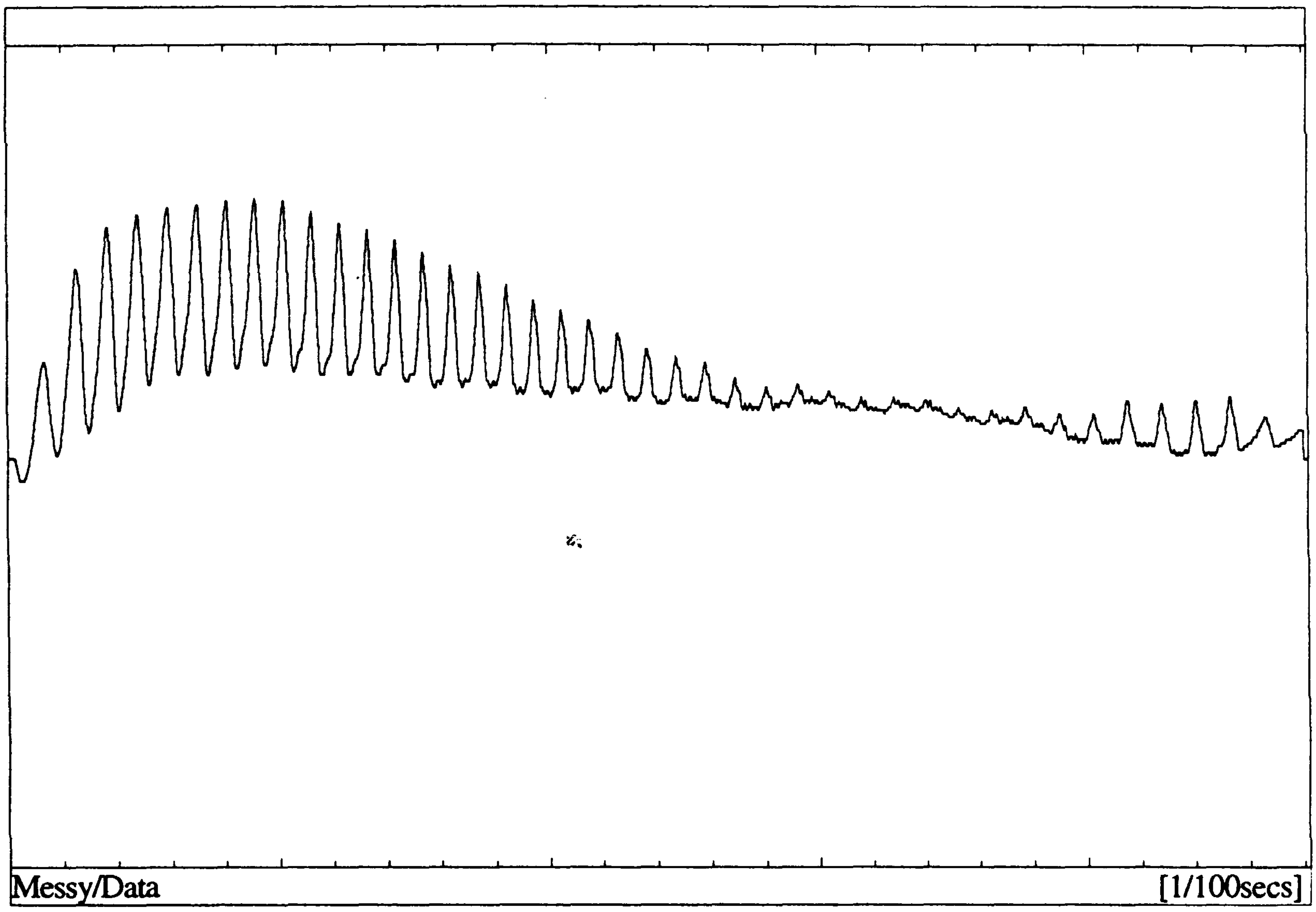

Figure 6. Chong 'messy' showing breathy onset and glottal wringing in the offset

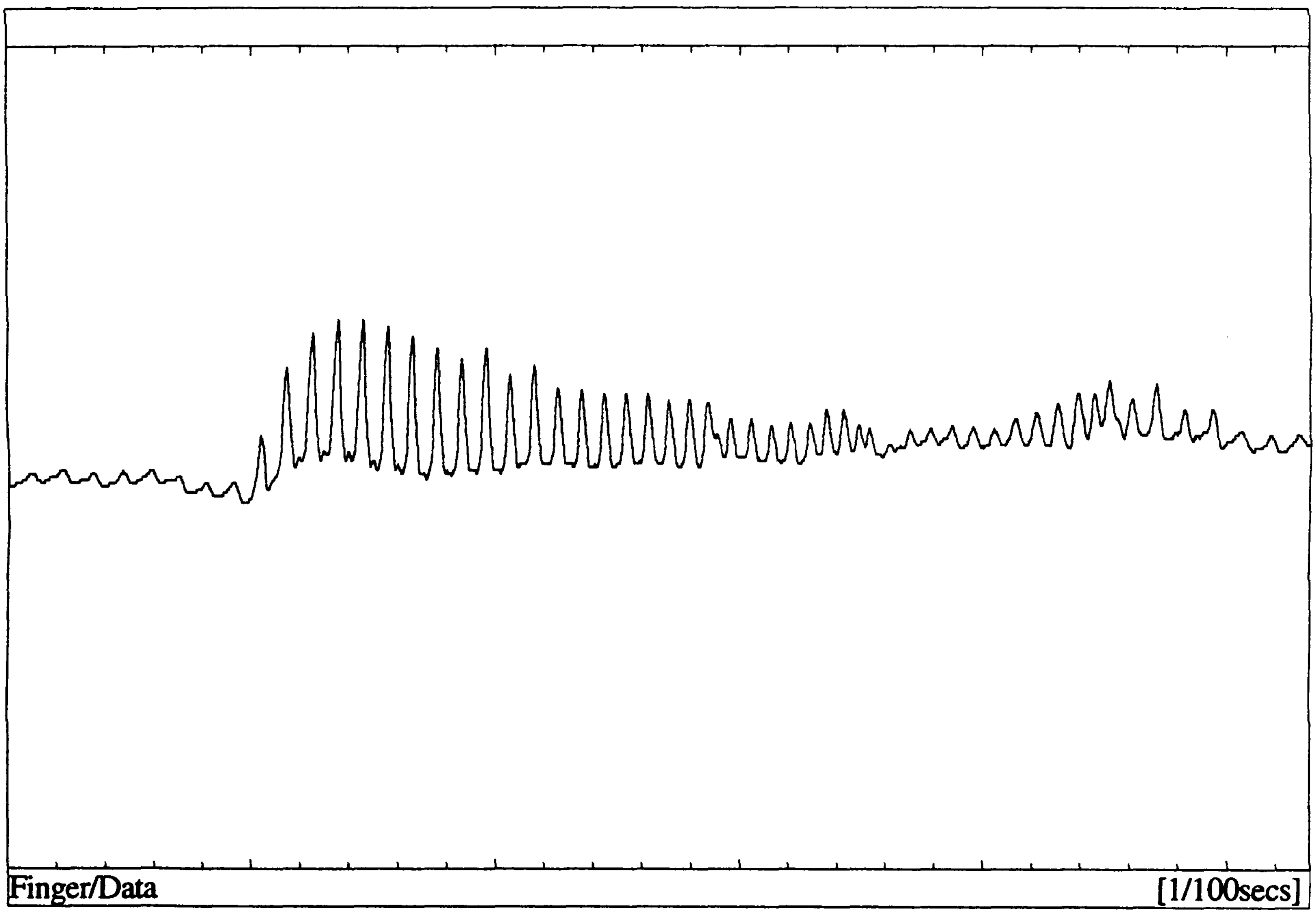

Figure 7. The second syllable in the word *ŋa̤a̤m* 'fingers' in Chong with breathy voice followed by dynamic glottal constriction and release

Figure 6 shows that the closed phase of the glottogram is maintained during the time of constriction and appears again after relaxing the constriction. Figure 7 shows another example of R4 in a less spectacular manifestation. Still, the high amplitude airflow wave at the beginning with symmetrical vibrations about the wave peaks is a certain indication of breathy voice. This effect comes from the lax and flaccid vocal fold tissue itself not being asymmetric.

4. Conclusion

These glottograms certainly support the claim in L-Thongkum of four "registers". They also reveal the strangling and constrictive nature of the "wringing" gesture, which was already observed by L-Thongkum as a drop in pitch and intensity with irregularity of fundamental frequency. All these features would be consistent with the description here. It was very impressive to me to observe native speakers producing such rapidly changing voice qualities in connected speech. It is little wonder that their voices sometimes sound a trifle "harsh" to speakers of languages which lack the "industrial strength" voice qualities found in Chong.

REFERENCES

Diffloth, Gerard. 1974. Austro-Asiatic languages. *Encyclopedia Britannica,* 480-4.

Diffloth, Gerard. 1992. "Palaungic vowels in Mon-Khmer perspective." In *Mon-Khmer Studies in Honour of Harry Shorto,* ed. J. H. C. S. Davidson. London: School of African and Oriental Studies, 13-27.

Edmondson, Jerold A. and Li Shaoni. 1988. "Voice quality settings and pitch in the Bai language of Yunnan Province." (in Chinese as Ai Jierui he Li Shaoni. 1989. Yunnan Jianchuan Baiyu yinzhi he yindiao anliang). *Zhongyang Minzu Xueyuan Xuebao* 70-74.

Edmondson, Jerold A. and Li Shaoni. 1991. "Voice quality and inverse filtering in the Bai language of Yunnan Province." In *Proceedings of the International Conference on Yi-Burmese Languages*, ed. Dai Qingxia. Xichang, Sichuan, Aug. 1991.

Edmondson, Jerold A. and Li Shaoni. 1995. "Voice quality and voice quality change in the Bai language of Yunnan Province." Submitted to *LTBA*.

Fant, Gunnar and B. Sonesson. 1962. "Indirect studies of glottal cycles by synchronous inverse filtering and photoelectric glottography." *Speech Transmission Laboratory Quarterly Progress and Status Report*-4/1962. 1-3.

Fant, Gunnar. 1960. *The Acoustic Theory of Speech Production.* The Hague: Mouton.

Fant, Gunnar. 1961. "A new anti-resonance circuit for inverse filtering." *Speech Transmission Laboratory Quarterly Progress and Status Report*-4/1961. 1-6.

Headley, Robert K. 1985. "Proto-Pearic and the classification of Pearic." In *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt,* eds. S. Ratanakul, D. Thomas, and S. Premsrirat. pp. 429-78. Bangkok: Mahidol University.

Huffman, Franklin. 1985. "The Phonology of Chong." In *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt,* eds. S. Ratanakul, D. Thomas, and S. Premsrirat. pp. 355-88. Bangkok: Mahidol University.

L-Thongkum, Theraphan. 1992. "An Instrumental study of Chong register." In *Mon-Khmer Studies in Honour of Harry Shorto,* ed. J. H. C. S. Davidson. pp. 141-60. London: School of African and Oriental Studies.

Laver, John. 1980. *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.

Lindqvist, J. 1965. "Studies of the voice source by means of inverse filtering." *Speech Transmission Laboratory Quarterly Progress and Status Report-* 2/1965.8-13.

Received: June 1995

Program in Linguistics
University of Texas at Arlington
Arlington, TX 76019-0559
USA
e-mail: jerry@ling.uta.edu