A Measurement from Electroglottography: DECPA, and its Application in Prosody

Alexis Michaud

Laboratoire Phonétique et Phonologie (UMR 7018) CNRS/ Sorbonne Nouvelle, Paris
Alexis.Michaud@univ-paris3.fr

Abstract

The present study, drawing on recent research on the use of the derivative of electroglottographic signals, applies to the study of intonation a parameter called DECPA, "Derivative-EGG Closure Peak Amplitude": the amplitude of the peak on the derivative of the EGG signal at glottal closure. A pilot study comparing DECPA, F_0, global intensity and open quotient (in data from tone languages) suggests that DECPA correlates with pragmatic emphasis *(accent)*. The uses and limitations of this measurement are discussed in relation to the general issue of the measurement of harmonic richness.

1. Introduction

1.1. The need for measurements of harmonic richness

It has long been known that vocal effort and perceived loudness are complex phenomena that measurements of global intensity alone cannot capture. The linguistic importance of the spectral slope has been shown for American English and Dutch [19]. As to the linguistic role of spectral tilt, the authors offer a clear-cut interpretation: high-frequency emphasis is a correlate of lexical *stress*, whereas variations in fundamental frequency (F_0) correlate with (intonative) *accent*, also called "focus target", "sentence accent". But many observations contradict this dichotomy. In Swedish [9], high-frequency emphasis was observed under focus, i.e. a pragmatic phenomenon. Fant *et al.* report spectral differences under lexical stress, but note that the strengthening of high frequencies is stronger under "emphatic stress" [8:623]. Similarly, [10] uses spectral slope measurements for the detection of what he calls "focus", "focal accent", "accent", "prominence", i.e. phenomena that correspond to [19]'s *accent*. In English data, an increase in energy in the higher frequencies was found under emphasis [1]. Thus, the linguistic importance of harmonic richness is not restricted to the marking of lexical stress, but is a major cue of the supralexical phenomena referred to in the literature as "accent", "focus", "prominence", "emphatic stress".

There is at present no efficient vowel-independent method for measuring the glottal flow wave and the spectral slope, e.g. inverse filtering is vowel-dependent and "theory-dependent", resting on certain theoretical assumptions. The argument of the present paper is that an interesting and relatively segment-independent cue can be obtained from electroglottography (hereafter EGG). Though EGG does not provide an estimate of the air flow at the glottis, it is a precise (and noninvasive) measurement of the contact surface of the vocal folds, by means of a high-frequency, low-intensity current passed between two electrodes placed on the neck. The parameters calculated from EGG, however, still vary broadly from author to author, substantiating the critical view of those who point out that EGG does not amount to a measurement of glottal flow, and warn against its unwarranted uses (especially as a tool for medical diagnosis).

1.2. New uses of electroglottography

There are however studies (based on high-speed cinematography) showing that parameters corresponding precisely to physiological events can be extracted reliably from EGG, by using the *derivative* of EGG (hereafter DEGG) : "All of the EGG waveforms exhibit a steeper slope (implying a rapid change in vocal fold contact area) during the glottal closing phase than during the opening phase. This characteristic phenomenon of vocal fold vibration results in a sharp negative pulse in the DEGG waveform at or near the instant of maximum glottal closure" [4:2400] (see figure 1, and [12]). The present pilot study is devoted to this sharp pulse in the DEGG signal, which is arguably related to harmonic richness: "Derivative-EGG Closure Peak Amplitude", or *DECPA*, the amplitude of the peak on the DEGG curve that corresponds to glottal closure. This parameter, pointed out to us by N. Henrich, was first studied in the singing voice, in [5:131-135] (a line of research currently pursued in [2]). It is also mentioned in [7], but the authors do not consider DECPA in itself: they calculate the ratio between DECPA and the amplitude of the opening peak. The latter parameter has no known incidence on the resulting sound wave, and our own experiments, which included the measurement of opening peaks, did not bring out any regularity in opening peak amplitude, so it seems more promising to study DECPA separately.

1.3. What is the significance of DECPA?

It is known that greater speed of glottis closure results in a less steep spectral slope. DECPA is a measurement at one single point in time for each period: it corresponds to the highest speed in increase of vocal fold contact surface which is reached at the glottis-closure-instant.

DECPA is measurable even for relatively noisy EGG recordings (such as those used in section 3). It is well known that the closing peak is strongly marked on the DEGG signal (by contrast, the opening peak is much less salient, and its detection is impossible in some cases [11]). DECPA can vary by a ratio of 1 to 5 within the same utterance. Its absolute value is not meaningful by itself, as it depends on the recording level set for the EGG signal; what is discussed in the following section is *the evolution of this parameter in time, over the whole utterance*. DECPA is compared with F_0, RMS amplitude, and open quotient.

2. Speech materials

The linguistic data under investigation is from tone languages: Naxi (a Tibeto-Burman language), Mandarin Chinese, and Vietnamese. In these languages, the lexical tones leave a margin of variation, which is used extensively for intonational purposes, in particular reflecting information structure. Languages which thus combine lexical tones and "intonational accents" offer an interesting background for a study of the prosodic marking of informational prominence. In Vietnamese, statistics on 504 target syllables in carrier sentences showed that DECPA is significantly higher under emphasis (see figure 4; details are given in another paper presented at the present conference: "Glottalized and nonglottalized tones under emphasis: open quotient curves remain stable, F_0 curve is modified"). As for Naxi, the materials reported on here are "classical" question-and-answer pairs. The corpus, originally designed for French [6], was translated into Mandarin Chinese, and again into Naxi with the help of an informant, retaining only the utterances for which he found natural-sounding equivalents. During the recording session, the Chinese version was read out as a prompt. C-programs were developed for this study by VU NGOC Tuân (LIMSI-CNRS) to compute F_0, the open quotient (hereafter O_q), and DECPA.

Predictably, several parameters were found to show differences between prosodically "strong" ("accented") and prosodically "weak" syllables, most notably F_0, duration, glottis adduction, the consonant-to-vowel ratio, and formant position. The question of the relative importance of individual cues, and of the (notoriously complex) trading relations found between the various cues, is beyond the scope of this presentation. The study will center on the behavior of the Mid and Low tones under focus (Naxi possesses lexical High, Mid and Low tones, and a Rising contour tone).

3. Results

3.1. The Low tone. In some cases, a DECPA maximum is observed on "focus", a salience not detectable by using F_0, O_q and global acoustic intensity.

Three types of behavior are reported in [15]: (i) intonation overrides the L lexical tone: the resulting F_0 is in the range of a H or M tone (i.e. Low tone under *accent* is sometimes as high as a nonaccented syllable with lexical High tone); (ii) a strong pharyngeal constriction is exercised and the syllable is devoiced; (iii) in many cases, there is *an extra effort which is not signaled only (or even not primarily) by a change in F_0.* This last category will be examined here, pointing out examples in which the focus does not stand out in terms of F_0 or global intensity, whereas it is characterized by a maximum of DECPA.

Figure 2 plots the DECPA and intensity curves for utterance (1).

(1) nɯ˩ kʰɑ˩dy˩ tʰv˧ bɯ˧ lɑ˥ ?

2nd pers. sg.-Beijing-to/(go) out-go-*interrogative particule*

"Are you going to Beijing ?" (context: two people are sitting on a train; answer: *"No, I am going to Tianjin."*)

The corresponding F_0 and O_q curves are found in figure 3. O_q gives an indication on the "pressed" quality of voicing and vocal fold adduction [16]. Neither O_q nor F_0 present any special prominence on the word "Beijing", /kʰɑ˩dy˩/. F_0 on both syllables is lower than on the 2nd-person pronoun "you"; in terms of F_0 glissando, the verb, /tʰv˧/, is most salient. Neither of the two syllables of "Beijing" stands out in terms of duration either. As for acoustic intensity, measured from the audio, the most intense syllable is the 2nd-person pronoun, at the beginning of the utterance. By contrast, DECPA is at its maximum on the first syllable of "Beijing", /kʰɑ˩dy˩/, which has greatest pragmatic/ informational importance.

3.2 The Mid tone: essentially similar observations.

In utterance (3), the "focus"/"accented syllable" is "seven", /ʂə˞˧/. Its F_0 is lower than that of classifier /tɕɑ˥/. On the curve of intensity, /ʂə˞˧/ does not stand out more than the negation /mə˧wɑ˩/. But, as in example 1, high DECPA values are briefly reached on the focus, /ʂə˞˧/ (see figures 5 and 6).

(2) nɯ˩ hu˩mɯ˧kʰv˥ tʂʰwɑ˥ dʐɯ˩ gə˧ xo˧tʂʰə˧ ndzɯ˩ pu˥ tsʰɯ˩ lɑ˥ ?

2nd pers. sg.-evening-six-o'clock-*GEN.*-train-sit-*"gerund"*-arrive-*interrogative final particle*

"Did you arrive by the 6 o'clock train?"

(3) mə˧wɑ˩, ŋə˩ ʂə˞˧ dʐɯ˩ gə˧ tʂʰɯ˧ tɕɑ˥ ndzɯ˧ tsʰɯ˩ mə˩.

neg.+copula-1st pers. sg.-seven-o'clock-*GEN.-deict.-class.*-sit-*"gerundif" (undergoes ellipsis)*-arrive-*affirmative final partic.*

"No, I arrived by the 7 o'clock one."

3.3 The importance of phonation mechanism

The Naxi examples above are from the speech of a male informant (as are the Vietnamese data of figure 4). Complexities arise when studying DECPA from recordings of female subjects, revealing that one further parameter must be taken into account in interpreting DECPA. This point will be made here by studying data from a female Chinese speaker.

The data consists simply of the sentence *He wants this pen*, 他要那只笔。 ta1 yao4 na4 zhi1 bi3 (*3rd pers. sg.*-want, need-*deictic-classifier*-pen) The two realizations in figures 7-8 correspond to the following interpretations: He wants this pen/ He wants this pen, with accent on different syllables. In figure 7, the accented (or "focused") element is *pen*, bi3, which carries lexical tone 3: a low-falling-rising tone. It appears that, despite its relatively low F_0 and intensity [intensity is not plotted here], bi3 is the most salient in terms of DECPA, following the pattern observed in section 2. The DECPA valley in the middle of the syllable is due to the fact that, as the speaker reaches the very bottom of her range, phonation is almost interrupted. Figure 8, on the contrary, shows paradoxical results: DECPA is low on the accented/focused word, where high values were expected.

This can be accounted for in light of research by B. Roubeau, which distinguishes four *laryngeal mechanisms*, on a physiological basis [17,18,13]. This four-way distinction is distinct from the classification of phonation types in [3,14]. Roubeau shows that in speech as well as in singing, phonation can be classified into several "laryngeal registers", or "laryngeal mechanisms": "mechanism 0", or *vocal fry*; "mechanism I", the most frequent in speech by male subjects, used by female subjects at the lower end of their F_0 range; "mechanism II" (alto/mezzo-soprano/soprano voice), in which the vibrating portion of the vocal folds is smaller (due to compressed arytenoids); and mechanism III, used for the very highest frequencies. It is essential to take the *laryngeal mechanism* parameter into account in studies of DECPA, as the vibrating portion of the vocal folds is significantly different in these four mechanisms, e.g. DECPA in mechanism II is lower, and shows much less variation, than in mechanism I (a physiological explanation is given in [17,18]). In our Chinese data, the speaker realizes high pitches in mechanism II, as female speakers are known to do [17]; low pitches in mechanism 1; and extra-low pitches, as in the lowest part of tone 3, in mechanism 0 (see annotation to figures 7-8). Three laryngeal registers are thus used in the utterance plotted in figure 8: register 2 during the first three syllables, which have high pitch (DECPA values: medium); register 1 during the fourth syllable, which has lower F_0 (DECPA values: high); register 0 during the last syllable, which is laryngealized throughout (DECPA values: very low), as evidenced by the very irregular glottal cycles. In figure 7, syllables 1 to 4 are in mechanism 2; syllable 5, which has higher pitch than in figure 8, is in mechanism 0 only in the middle (with very low DECPA values), and in mechanism 1 at the beginning and end, where the F_0 values are higher. Similar observations were made with a female Naxi speaker.

4 General discussion and conclusion : Use and limitations of DECPA in prosodic studies

The limitations of DECPA are:

- Direct comparison can be made neither across speakers nor across recording sessions, as DECPA varies with input level.
- Comparison is only possible within the same laryngeal mechanism. This makes it necessary to determine which mechanism is being used. This is important information, for which no automatic detection procedures have been proposed as yet [13].
- If there is a double closing peak, which is frequent, no principled answer can be given as yet on which is the “true” DECPA: the first, the higher of the two, or some value computed from the two peaks. (This topic is currently investigated by N. Henrich and C. Gendrot by means of high-speed cinematography.)
- Quantifying the relationship between DECPA and spectral slope is complex. Points of reference are missing; e.g., a comparison of spectral slices and DECPA values at different points in the course of a syllable does not offer a reliable basis because of the constantly changing vocal tract shape (and formant frequencies). More generally, it is unlikely that this relationship can be expressed in a universally valid mathematical formula, as DECPA, the maximum speed at closure, is only one of several glottal source parameters.

With these reservations, the proposed use of DECPA as part of an experimental setup in linguistic studies of stress, accent and vocal effort is to detect the points of DECPA prominence within an utterance, or across items within a recording session. From our observations, it appears that the linguistic phenomena of *accent* is better correlated with the DECPA maxima reached (even very briefly) than with DECPA values averaged over syllables. In conclusion, DECPA does not actually amount to the much-wanted measurement of spectral slope, but represents a significant cue in the case of phonation mechanism I.

Resource sharing - Hypertext link

The signals of the examples are currently available at www.cavi.univ-paris3.fr/ilpga/ED/student/stam/index.html. Requests for the C programs for EGG analysis should be sent to Tuan.Vu-Ngoc@limsi.fr.

Acknowledgments

Many thanks to N. Henrich (LAM/CNRS) and T. Vu Ngoc (LIMSI/CNRS) for providing guidance and software. Special thanks also to J. Vaissière, M. Castellengo, K. Kohler, and A. Rialland, C. Mooshammer and C. Gendrot; to Pr. Sun Hongkai and Huang Xing for the invitation to China; to Pr. Kong Jiangping for use of the facilities at his laboratory in Beijing; and to the Naxi and Vietnamese informants and friends. *Financial support was given by the Fondation de France and by the Conseil Scientifique of University Paris 3.*

5 References

[1] Campbell, W.N., 1995. Loudness, spectral tilt, and perceived prominence in dialogues. XIIIth International Congress of Phonetic Sciences, Stockholm*:* 676-679.

[2] Castellengo, M.; Chuberre, B.; Henrich, N. (submitted). The "voix mixte" register: a way of managing the transition between the two main laryngeal mechanisms. International Symposium on Musical Acoustics 2004, Nara, Japan

[3] Catford, J.C., 1964. Phonation types: The classification of some laryngeal components of speech production. *In Honour of Daniel Jones*. D. Abercrombie et al. London, Longmans, Green and Co.: 26-37.

[4] Childers, D.G.; Lee, C.K., 1991. Vocal quality factors: Analysis, synthesis and perception. *Journal of the Acoustical Society of America* 90(5): 2394-2410.

[5] Chuberre, B., 2000. *Les registres et passages dans la voix chantée*, Ph.D., Université de Nantes (France).

[6] Delais-Roussarie, E.; Rialland, A.; Doetjes, J*., et al.*, 2002. The Prosody of Post-focus sequences in French. Speech Prosody, Aix-en-Provence*:* 239-242.

[7] Dromey, C.; Stathopoulos, E.T.; Sapienza, C.M., 1992. Glottal airflow and electroglottographic measures of vocal function at multiple intensities. *Journal of Voice* 6(1): 44-54.

[8] Fant, G.; Kruckenberg, A., 1995. The voice source in prosody. XIIIth International Congress of Phonetic Sciences, Stockholm*:* 622-625.

[9] Gobl, C., 1988. Voice source dynamics in connected speech. *STL-QPSR* 1: 123-159.

[10] Heldner, M., 2003. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* 31(1): 39-62.

[11] Henrich, N., 2001. *Etude de la source glottique en voix parlée et chantée*, Ph.D., Université Paris 6 (France).

[12] Henrich, N.; d'Alessandro, C.; Castellengo, M*., et al.*, accepted. On the use of the derivative of electroglottographic signals for characterization of non-pathological voice phonation. *Journal of the Acoustical Society of America*.

[13] Henrich, N.; Roubeau, B.; Castellengo, M., 2003. On the use of electroglottography for characterisation of the laryngeal mechanisms. Stockholm Music Acoustics Conference, Sweden

[14] Laver, J., 1980. *The phonetic description of voice quality*. Cambridge, Cambridge University Press.

[15] Michaud, A., in press. Focus et tons lexicaux: de la variété des paramètres prosodiques. *Fonctions et moyens d'expression de la focalisation à travers les langues*. A. Lacheret-Dujour and J. François. Louvain, Peeters.

[16] Rothenberg, M.; Mahshie, J.J., 1988. Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research* 31: 338-51.

[17] Roubeau, B.; Chevrie-Muller, C.; Arabia-Guidet, C., 1987. Electroglottographic study of the changes of voice registers. *Folia Phoniatrica* 39: 280-89.

[18] Roubeau, B.; Chevrie-Muller, C.; Arabia-Guidet, C., 1991. Control of laryngeal vibration in register change. *Vocal Fold Physiology: Acoustic, perceptual, and physiological aspects of voice mechanisms*. J. Gauffin and B. Hammarberg. San Diego, Singular Publishing Group: 279-86.

[19] Sluijter, A.; Van Heuven, V.J.; Pacilly, J.J.A., 1997. Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America* 101: 503-513.

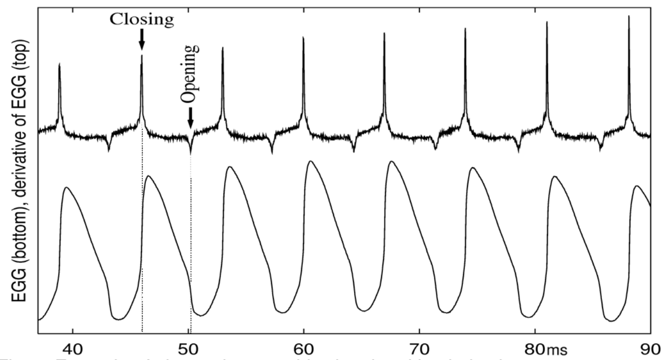

Fig. 1. Example of electroglottographic signal and its derivative (data: Vietnamese)

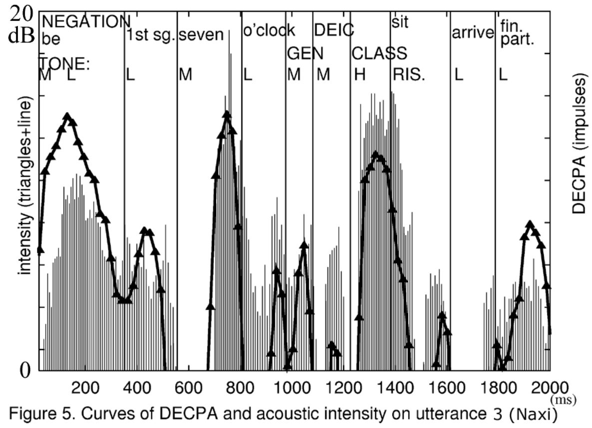

Figure 5. Curves of DECPA and acoustic intensity on utterance 3 (Naxi)

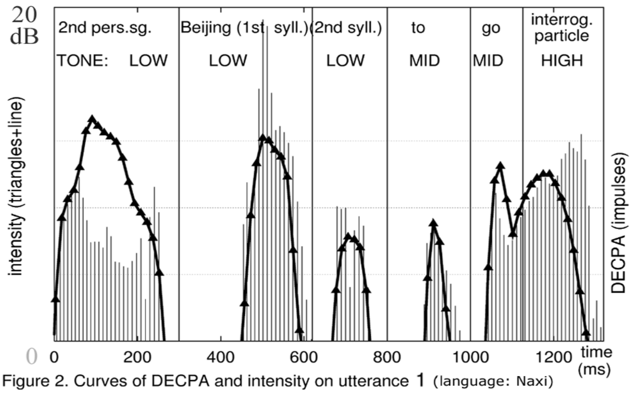

Figure 2. Curves of DECPA and intensity on utterance 1 (language: Naxi)

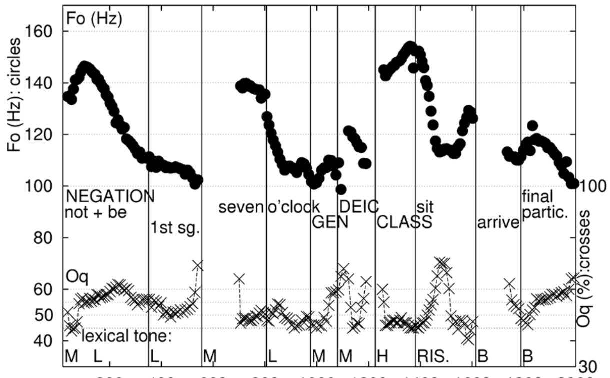

Figure 6. Curves of fundamental frequency and open quotient on utterance 3

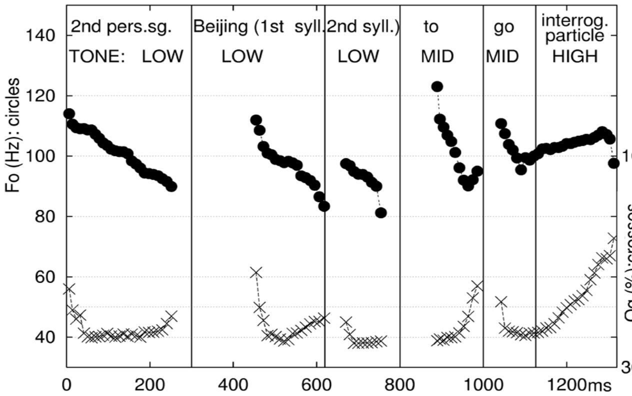

Figure 3. Curves of fundamental frequency and open quotient on utterance 1.

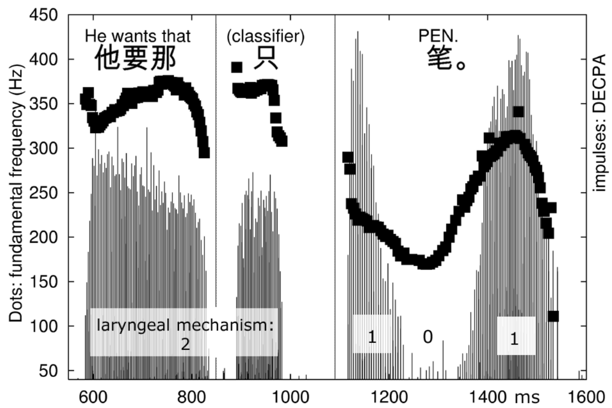

Figure 7. Accent on syllable 'pen' (tone 3). (Mandarin Chinese)

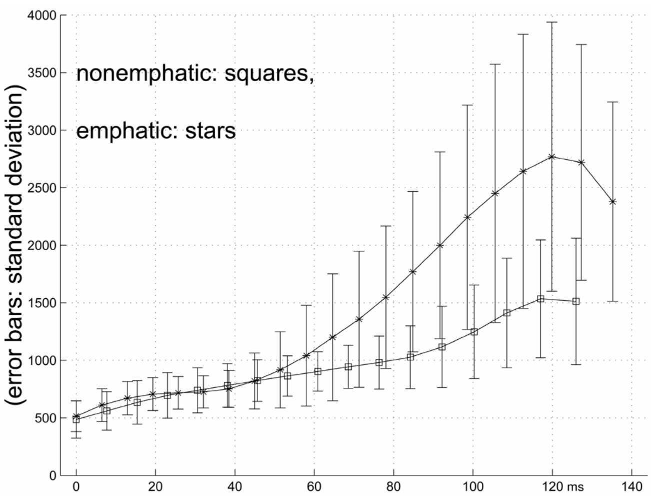

Figure 4. Averaged DECPA curves for Vietnamese tone 8

Contrast of nonemphatic and emphatic reading. 1 (male) speaker, 82 syll.

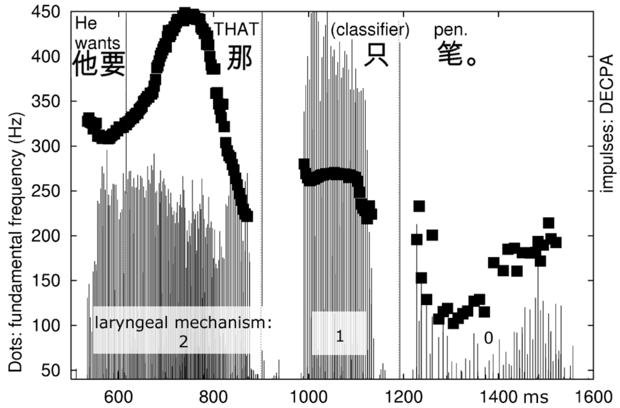

Figure 8. Accent on deictic (Falling tone: tone 4). (Mandarin Chinese)