# A SEMI-AUTOMATED LX-BASED METHOD FOR THE MEASUREMENT OF VOICE ONSET TIME

*Krzysztof Marasek*
*University of Stuttgart, Institute of Natural Language Processing*
*Experimental Phonetics, Stuttgart, Germany*

## ABSTRACT

The main goal of this study is to establish a reliable automated method of Voice Onset Time (VOT) estimation. It is shown, that VOT measurements can reliably and accurately be performed by combining the information about stop release from acoustic signal with the information about voicing initiation derived from the laryngographic signal. This task can be performed automatically.

## PREFACE

VOT, defined as the time difference between stop release of a plosive and the start of vocalisation of the following vowel, is a common parameter in the investigation of speech and language disorders [2]. The proposed method uses two-channels recording of the speech signal. The Laryngograph was used to monitor the activity of the vocal folds (Lx signal) with the acoustic speech signal simultaneously recorded on the second channel (Sx signal). The starting time instant of vocal fold vibration is based on the Lx signal analysis, while the closure release impulse is found in the Sx signal.

## THE Lx SIGNAL

The Laryngograph [3] enables direct measurement of vocal fold activity. Thanks its non-invasive measure method (laryngeal conductance measured by pair of electrodes situated on the neck, both-sides of the cricoid cartilage) and mostly high SNR ratio, the laryngograph is often used as a reference signal for pitch period determination. Nonetheless, the output from laryngograph (Lx) is not free from problems: it is influenced by vertical movements of the larynx (so called Gx signal) and it does not match some movements of the vocal cords. For some speakers the Lx regristration may even fail temporarily. The amplitude of the Lx depends on speech loudness. However, as it was confirmed in [1] the Lx signal matches exactly the vocal fundamental frequency. The Lx signal is also used in more detailed analysis of vocal folds activity. Of special interest is the use of the Lx signal to differentiate pathological modes of phonation. To achieve that, an undistorted form of the Lx should be used, i.e. the influence of the Gx should be eliminated, but the distortion of the shape of the Lx waveform shuld be avoided at the same time.

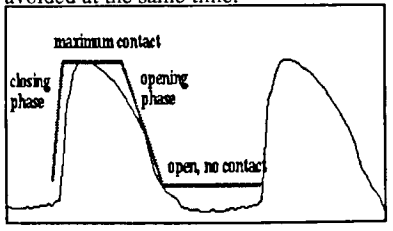


*Figure 1. Phases of the Lx signal.*

The individual pitch periods of speech signal are determinated in the Lx signal through position of the peak change of the laryngeal conductance, which is adequate to the instant of glottal closure. The changes of the conductance during one pitch period are presented on Fig.1. When the glottis is open, the conductance is low and flat. During closing phase, the conductance rises steeply and remains high during closure. Then, during the opening phase of the vocal folds the conductance is falling, but not so steeply as during the closing phase. The position of the maximum change of the conductance is determinated by zero-crossing and/or by thresholding of differentiated Lx signal. Previous to the use of this method it is necessary to pass the Lx signal through low-pass filter to avoid the influence of the Gx signal. The vertical movement of the larynx may be fast, so it is not easy to determine cut-off frequency of the filter. As it was pointed by Baken [1] such filtering may strongly influence the shape of laryngographic waveform, making it unsuitable for



further analysis. Hess and Indefrey [4] proposed more sophisticated algorithm (with very good temporal resolution), but it also needs filtering of the Lx signal and fails in case of rapid vertical movements of the larynx.

The proposed algorithm works on the raw, unfiltered Lx data. The algorithm may be divided into 3 steps:

1. The markers are set at the positions of local maxima. Markers are set only when the local maximum occurs after given time (thresholding in the time domain) and next samples differs significantly in amplitude to the maximum (thresholding in the amplitude). Then the positions of local minima are found, also with tresholding in time and amplitude domains. Further analysis is based on the pair of markers: maximum-minimum. The temporal difference between them may be used as pitch period estimate. The positions and amplitudes of maxima and minima are compared to their neighbours and, when they are significantly weaker and shorter, they are attached to stronger pairs (such situation occurs in creak-like or laryngalized phonation).
2. The parameters of whole record of the data are taken into account in the second pass of the analysis. The pairs of markers which occur in isolation or in very short train of markers (<4) are recognised as an error and removed if they occur between long unvoiced segments of speech (at least 200 ms). Differences in the pitch periods length are analysed and if the length quotient is greater than 5, an algorithm tries to recover the min-max pairs from the original signal. Thereafter the begining and the end of each voiced segment is marked. In every voiced segment each maximum-minimum pair is again analysed to find disturbances from mean length and mean relative amplitude within the pitch periods. If deviations are greater then given threshold, the weaker pairs are connected to the stronger ones (if a resulting pair is not too long regarding mean pitch length). The min-max pairs segment the Lx signal according to pitch period length.
3. The glottal shape encoded in the Lx waveform is established on the third pass of analysis. Based on minimum-maximum pairs the time instants of the opening, the open, the closing and the close phases of the glottis are found. To find the time instant of the approximated begining of the closing phase the 3-point smoothed Lx waveform is analysed and the point of maximum of the first difference is chosen. The starting point of opening phase is more difficult to find, especially when rapid movements of the larynx occur. It is assumed, that the opening starts at the same level of conductance as at the begining of the closing phase, so the next start of the opening (point 5 on Fig.2) is found as the crossing point of Lx waveform with straight line connected to the closing phase markers (dotted line between points 2 and 5 in Fig.2).

Every period of Lx signal is than described as presented on Fig.2.

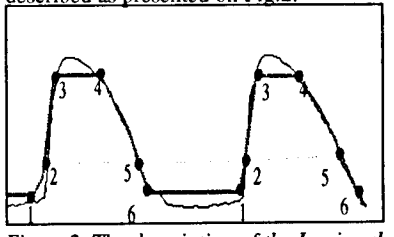


*Figure 2. The description of the Lx signal using straight lines.*

The shape is described using six idealized straight lines and deviations from those lines are used as indices for the signal classification.

## THE STOP RELEASE

The impulse characteristic the stop release has to be found in speech channel of the record.

The direction of search for stop release depends on the form of the speech signal after the start of the vocalisation. In fact, the vocalised friction phases of plosives contains only low-frequency component. Thus, the zero-crossing (ZC) rate for every pitch segment is very low, distinguishing it from following vowel segments (after F2 release). The decision is made based on the ZC rate for the first 10 periods of the speech signal. If ZC rate is low (and its variation is also low) it is assumed, that the VOT will be negative. The closure impulse for the negative

VOT forms a short (and rather weak) noise-like burst on the top of speech signal (Fig.3). In order to find this impulse the differentiated speech signal is consulted, but only on its positive parts. The segment with the highest ZC-rate points the period with the closure (short noise burst). Within the founded segment the greatest jump in the amplitude of the speech signal points to the release of the impulse. Additional conditions on length and strength of the burst prevent accidental determination of the VOT.

The search for the stop release in positive direction is based on the difference in the signal energy between the silence (occuring before closure) and other phases of the plosive. In fact, between the begining of the stop release and the start of the following vowel (i.e. in the burst phase) some noise is present, thus short time energy shows a rapid step indicating the start of the plosive (given appropiate SNR and the initial position of the plosive). Thereafter, in a window where the energy changes most rapidly, the time index of the sample with the biggest difference is determinated as the starting point of the closure release. As a validation, the energy between segments before and after the stop release is compared (the following one should be bigger than the preceding one).

## RESULTS

The method was tested on recordings done using the Laryngograph processor (Lx) and small electret microphone (Sx). The recordings included 3-4 logotomised words (like : /baba/, /papa/, /gaga/). The method was tested not only on normal speakers with modal voices, but also on patients (with neurological disorders) showing some voice disoders (breathy creaky voice). The recordings of the patients have substantially lower quality as the control ones, especially regarding the SNRs. The VOT was measured only in the initial position. The results are summarised in Table I.

## DISCUSSION

As can be seen from Table I the results, althought quite good for so complicated signals, are not fully satisfactory regarding the percentage error. The most errors were caused not by troubles in the perfect localisation of the closure release impulse, but rather by the imprecise localisation of the start of vocal fold vibration. Within the vocalic segment, the first one or two periods of vibration are destroyed, their amplitude and duration is irregular and non-stationary (see Fig.4). To overcome this, a kind of soft-tresholding (the parameters are used with additional weights) in the amplitude and the time domain was used to find the begining of the Lx-vibration. This method was quite sucessfull for control speakers (the one major error within this group was caused by intentionaly unnatural, very long negative VOT) but failed for other groups of speakers, whose speech was very slow and quiet. It was observed, that for so quietly speaking persons, the Lx signal was distored or even lost for some moments. The Lx signal exceeded also the permited range of the A/D converter due to rapid movements of the larynx (swallowing). The minor errors (smaller than 1 ms, typically about 0,5 ms) are caused by distrurbances in location of the closure impulse. The overall description of the Lx signal, however, performed well and the single periods were precisely located.

## CONLUSIONS

It was shown that VOT measurements can reliably and accurately be performed by combining the information about stop release from acoustic signal with the information about voicing initiation derived from the laryngographic signal [2] and that task can be performed automatically.

## References.

[1]BakenR.J.(1992),Electroglottography, J. of Voice, Vol.6, No.2, 98-110.

[2] Blumstein S., Cooper W., Goodglas H., Stalender S., Gottlieb J. (1980), Production Deficits in Aphasia: A Voice-Onset Time Analysis, Brain &Language 9:153-170.

[3] Fourcin A.J. (1993), Normal and pathological speech: phonetic, acoustic and laryngographic aspects, in: Singh W., Soutar D., Functional Surgery of the Larynx and Pharynx, Butterworth Heinemann.

[4] Hess W.J.(1992), Pitch and Voicing Determination, in: Furui, Sondhi M.M. (eds.) Advances in Speech Signal Processing, Marcel Dekker Inc., 3-49.



*Table I. The results of VOT measurements for different groups of speakers. The numbers given in square brackets [] describe segments, where VOT was found, but in reality there was no plosive at the begining of the voiced segment.*

| Group of voices | Number of VOTs | No. of small errors (< 1ms) | No. of big errors (> 1ms) | No. of not or false recognised VOTs |
|---|---|---|---|---|
| control - modal voice | 15 | 3 | 1 | [1] |
| aphasia - modal voice | 12 | 4 | 3 | 1 |
| dysarthia - creaky voice | 12 | 5 | 3 | 2 |
| parkinsons - breathy voice | 12 | 4 | 3 | [1] |
| Σ | 51 | 16 | 10 | 6[2] |
| % errors | | 31 | 19 | 11(5) |

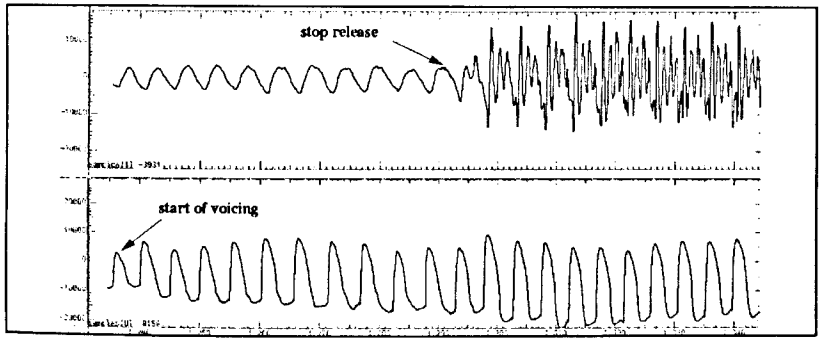


*Figure 3. An example of the negative VOT.*

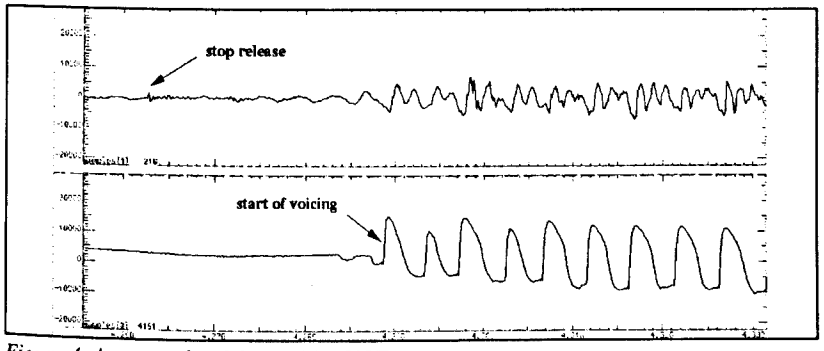


*Figure 4. An example of the positive VOT.*