Fo curves – smooth, seamless yet pulsed?

Björn Lindblom
Department of Linguistics, Stockholm University, 10691 Stockholm, Sweden

Abstract

This paper presents pilot data and analyses suggesting that Central Swedish F0 contours can be described with good numerical accuracy as sequences of syllabic overlapping quasi-Gaussian pulses. Each such building block is temporally aligned with a given syllable. It is specified by its 'amplitude' (peak frequency) and the temporal extent of its rise and fall slopes. According to the proposed account an F0 curve is seen as the sum of the bell-shaped pulses and a declination ramp.

In revisiting classic publications on respiratory and phonatory speech processes one notes that, although Stetson's 'chest pulse' notion is rejected by most phoneticians today, one nonetheless finds a great deal of evidence for increased P_s and EMG activity in stressed syllables (language examples: English and Swedish). Curiously, some of that evidence comes from the work by Ladefoged, one of Stetson's fiercest critics.

Are the results of the present modeling merely a fluke outcome of the curve fitting exercise, or do they tell us something significant about the motor control of prosodic phenomena in Swedish?. The following facts lend preliminary support to the latter possibility. ***1****. Syllables with main and secondary lexical stress show larger peaks than syllables with weak or no lexical stress.* ***2****. Pulses overlap in time thereby often losing their original discrete nature and hiding a possible pulsatile origin of F0 characteristics.* ***3****. The bell-shaped pulses contribute a rise-fall component to the output F0 trace and seem well suited to handling the correlates of the tonal accents. As expected the main grave-acute difference was found to be one of timing.* ***4.*** *One pulse per syllable was sufficient. It occurred within the time frame of the syllable. However the surface location of an F0 peak was sometimes temporally shifted away from its associate syllable (as seen in 4, 40, 400 and 32, 320, 3200 sequences). This phenomenon turned out to be successfully predicted by the syllable-aligned superposition model.*

The notion of 'chest pulse'

Stetson and Ladefoged

In 1951 Stetson proposed that there is an identifiable ballistic '*chest pulse*' for each syllable. He based this claim on measurements of rib cage movements, air pressure in the trachea and the lungs and on some EMG records. Partly, we may presume, he also used his intuition.

Ladefoged's *Three areas of experimental phonetics* (1967) contains a summary of research that refutes Stetson's hypothesis. Ladefoged found Stetson's work technically unreliable and was unable to identify chest pulses in his own experimental records: "*As might be expected from our electromyographic studies, we could find no correlation between subglottal pressure and syllables. It is clear, not only from the records we have been discussing here, but also from hundreds of others, that Stetson (1951) is wrong in claiming that there is a relation between respiratory activity and the syllable …*" (Ladefoged 1967, p 46-47).

He continues: "*Thus figures 17 and 18, which show even the small variations in subglottal pressure due to each opening and closing of the vocal cords, do* ***not show peaks*** *of subglottal pressure* ***which can be correlated with the individual unstressed syllables****.*

We can only presume that, in so far as Stetson's records sometimes indicate different results, it must have been because his subjects were not talking in a normal conversational tone. It seems probable that they were talking more loudly, slowly and distinctly than is customary.

In our opinion ***there is certainly insufficient basis for a chest pulse theory of the syllable in normal speech****.*"

Subsequently the field of phonetics has accepted Ladefoged's criticism (cf Lehiste

1970:109). Currently it would no doubt take a dim view of any attempt to revive the chest pulse notion.

Let us turn to Ladefoged's Figure 18 which compares subglottal pressure for statements and questions and for disyllabic noun-verb pairs distinguished by stress placement: *'digest'*, '*torment'*, '*pervert'*, '*survey'*. We here focus on the strong-weak vs weak-strong patterns of the nouns and verbs. Ladefoged notes that there is only one P_s peak per word. Moreover, in nouns the peak occurs on the first vowel; in verbs on the second. "*It is apparent that* ***every stress is accompanied by an extra increase of subglottal pressure***" (p 46).

We should also examine more closely some of his EMG findings, for instance his Figure 8 which shows the activity of the (expiratory) internal intercostals during '*The old man doddered along the road*'. A single motor unit was captured and its instantaneous frequency was calculated in impulses per second. This parameter is plotted in Figure 1

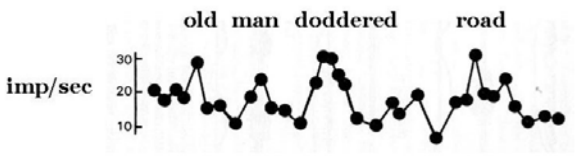

Figure 1. Firing rate of single internal intercostal motor unit as a function of time [adapted from Figure 8 in Ladefoged 1967].

The stressed words are indicated at the top of Figure 1. It can be seen that the stresses are clearly associated with greater activity in the respiratory muscle under observation.

In two more recent publications Ladefoged resumed his EMG studies to replicate his findings (Ladefoged & Loeb 2002, Ladefoged 2005). Single unit measurements were made in the internal intercostals during '*The old man doddered along the road'*. A closely similar result was obtained. His overall conclusion (boldface ours): " ***stressed syllables*** *may or may not have a greater intensity or a higher pitch, and they may or may not have bursts of internal intercostals activity. However, they* ***always use greater respiratory energy***." (Ladefoged 2005:20).

Singers and actors

Sundberg et al (1995) contains a diagram of F0, sound level and esophageal pressure recorded as a professional bartone sings a descending scale with three tones on each scale tone and with emphasis on the first beat in each bar. The subglottal pressure trace reflects this prominence by a clear increase.

Another example is the oesophageal pressure variations of a speaker who recites "*Kung Karl, den unge hjälte, Han stod i rök och damm…...*" in a loud, clear and overarticulated style ('stage speech'). The listener perceives a strong prominence on *'Karl'*, *'ung'* and *'hjält'* and again the oesophageal pressure trace shows high values for those syllables.

From chest pulse to stress pulse?

Ladefoged's criticism of Stetson's methodology is well taken. And we have no reason to doubt the absence of measureable correlates of 'pulses' in Ladefoged's data, e g, no peaks of subglottal pressure which can be correlated with the individual unstressed syllables.

The curious thing though is that his own EMG and P_s measurements do demonstrate that some sort of 'pulse' (extra respiratory activity) is associated with the syllable. Would it be fair to say that Ladefoged's data show the existence of a 'chest pulse' ("greater respiratory energy") at least in stressed syllables in English? In other words a 'stress pulse'?

I believe that to be a reasonable interpretation. Below its implications for F0 analysis are explored.

A pilot experiment: Procedures

The data of this report come from a male (phonetically trained) speaker of Central Swedish (= yours truly). He was recorded producing words and short phrases selected as follows.

Speech samples and recordings

Utterance length: From monosyllabic up to five syllables. Prosodic patterns: Grammatically legal permutations of four prosodic syllable types: (a) *main stress* ('huvudtryck') and acute accent (in SAOB notation a '4' syllable); (b) main stress ('huvudtryck') and grave accent (a '3' syllable); (c) *secondary stress* ('starkt bitryck', a '2' syllable); (d) *weak or no stress* ('svagtryck', here denoted by '0'). Words with the grave accent were consistently compounds with their obligatory secondary stress.

The subject read pairs of words/phrases: a meaningful word or phrase (examples in Table 1) followed by the corresponding test word as a '*re-iterant*' form - a (nonsense) word that has the same prosodic pattern as the model word and that uses segmental content in a controlled way.

4	sup
40	fjäder
04	pannå
400	kritiker
040	kaninen
004	logistik
4000	serierna
0400	botaniker
0040	intressera
0004	professionell
44	röd bil
444	Bengt går fort
440	Ulf läser
404	bra musik
044	Katrin vet
4004	Ted muckar gräl
40004	Per leker polis

32	kråkspark
320	brevlåda
302	motorbåt
032	miljövård
3200	busgrabbarna
3020	Londongatan
3002	väderprognos
0320	rabattkortet
0302	polispiket
0032	pyramidform

Table 1. Examples of prosodic patterns investigated. The numerical notation of stress and tonal accents is similar to that of SAOB (Elert 1970).

At least five tokens of each word or phrase were recorded. A later check of the recordings indicated that the subject had successfully maintained constant normal vocal effort and a pattern of uniform rhythm and intonation for the majority of speech samples.

kritiker	[ˈlɑ:nanan]
MacMillan	[laˈmɑ:nan]
aladåb	[lalaˈnɑ:n]

Table 2. Examples of model words and their 're-iterant' variants.

Re-iterant items were constructed by replacing single consonants and consonant sequences by [l]. [m] or [n] and by substituting [ɑ:] for vowels with main or secondary stress and [a] for other vowels. The use of sonorant, less constricted consonants reduces microprosodic perturbations of the F0 curve. Examples of *re-iterant* forms are given in Table 2.

F0 tracking & deriving average contours

Waveform files (sampling rate = 44100 Hz) were opened in Wavesurfer 1.7.5 for F0 analysis using the ESPS autocorrelation option. The output table contained F0 sampled every 10 ms. Time locations of vowel onsets and offsets were also tabulated.

For each test item repetitions were lined up in EXCEL side by side for synchronization and derivation of an average contour. As line up points vowel onsets were used. The averaging was performed across tokens once a syllable-by-syllable synchronization had been achieved. The alignment procedure was to locally shift the time scale of each syllable so as to match its vowel onset to the corresponding vowel onset of a reference token. This meant erasing or interpolating between values. These time shifts were generally small and were preferably made at points of minimum rate of F0 change. The F0 curves of this report are all derived by means of this procedure.

Results

Declination component

The F0 traces consistently showed a declination effect. To fit pulses to the Fo contours more conveniently I looked for a method to remove the declination component.

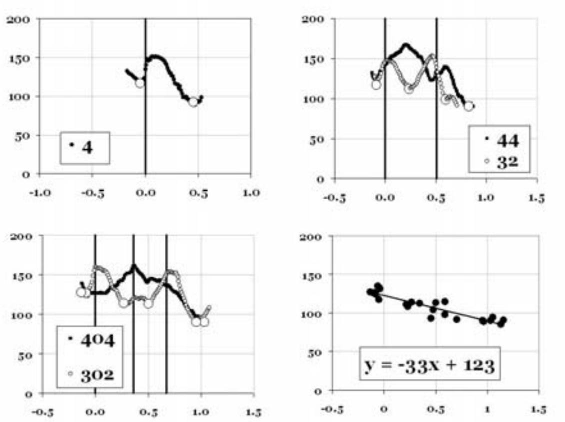

Figure 2. The declination ramp. Average F0 curves for selected test items. Vowel onsets are indicated by the vertical lines. Large open circles show points that were measured to quantify the declination effect. Time zero occurs at the first vowel onset. Bottom right: When these points are plotted on the same diagram they form a linear cluster with a falling slope of 33 Hz per second.

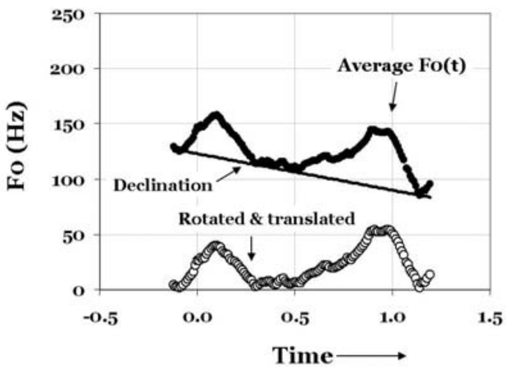

Figure 3. Data preparation. Removal of declination for curve fitting.

In the further processing of the data the declination component was removed by rotating the pattern so as to make it horizontal and translating it so as to place its minimum points on the zero line. Figure 3 summarizes the procedure.

The pulse

For a mathematical description of the bumps of the rotated and translated curves (cf Figure 3) I chose the Gauss curve:

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

letting x equal time and redefining μ as the time location of the pulse peak and using σ as a measure of the temporal extent of the rise and fall segments. In most cases the σ values for the rise and the fall were not equal implying asymmetrical pulse shapes.

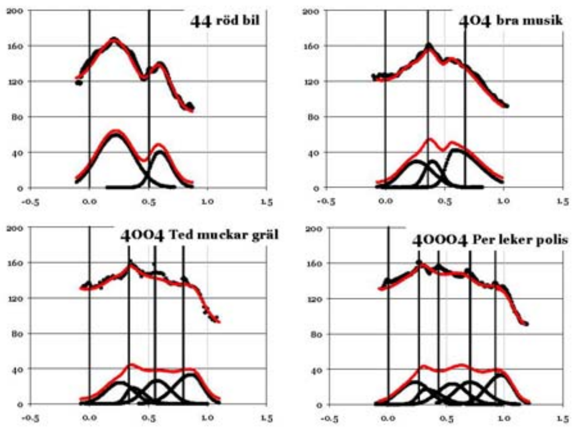

Figure 4. Phrases containing syllables with accent I. At the top of each panel: F0 contour (black) and the calculated curve (red). Below: Pulses individually (black) and summed (red).

Modeling the tonal accents

Figure 4 presents four phrases whose initial and final syllables have main stress and the acute accent. The top of the panels compares the observed F0 contour with a calculated curve. As can be seen there is a rather close match.

The bottom part shows the individual pulses (black) and a curve representing their sum (red). This curve was translated and rotated (=putting the declination back in) to derive the calculated trace.

These phrases were spoken in a 'legato' fashion. Note for the 44 pattern (top left panel) that there is no juncture between 'röd' and 'bil'. Nonetheless the F0 trace exhibits a valley in between the peaks which the pulse model handles satisfactorily.

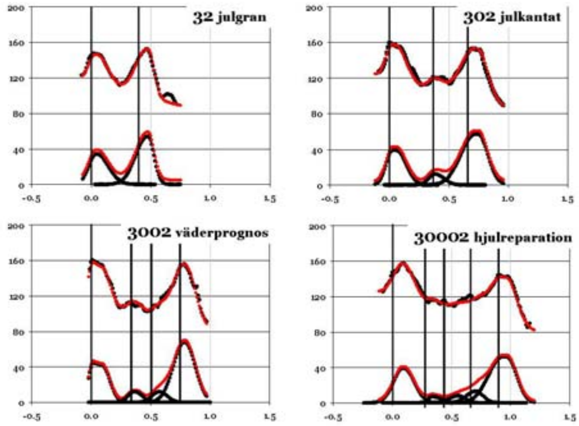

Figure 5. Phrases containing syllables with accent II.

Figure 5 is analogous to Figure 4. It presents four words with the compound 3(-)2 pattern. Initial syllables have main stress and the grave accent. Final syllables have secondary stress. Again we see a close match.

Figures 4 and 5 are representative of all test items analyzed.

Figure 6 applies the pulse analysis of two phrases differing in the placement of contrastive stress.

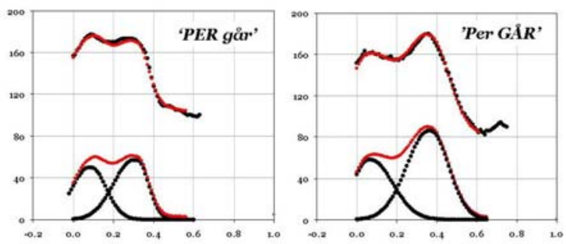

Figure 6. Two '44' sequences with emphatic stress on the first and the second syllable respectively.

Comparing grave and acute patterns

Figure 7 presents a comparison of the phrase '*grå ben*' (two adjacent acute accents) and the 32 compund '*gråben*'. Again it is noteworthy that the 44 pattern was spoken 'legato' without a juncture. Nonetheless there is a dip in F0 between the peaks - a fact which is compatible with the present pulse analysis.

The grave/acute contrast is here mainly linked to the timing of the first stress pulse and to the faster fall of the grave pulse.

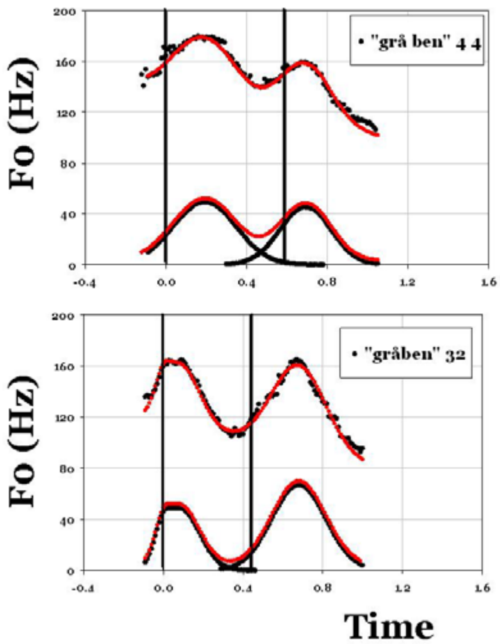

Figure 7. A 'minimal pair' illustrating the similarity between a 44 phrase and a 32 pattern.

Riad (1998) has proposed that historically the emergence of Scandinavian tonal accents is linked to a situation with two adjacent stresses and a need to resolve this 'stress clash' ('betonginskrock'). It is interesting to place his discussion of tone accent origins in the context of the pulse model. In a sense the top 44 phrase in Figure 7 represents a stress clash. The reorganization needed to go from the 44 to the 32 pattern is a moderate one: a matter of a small timing shift and possibly a rate change in the fall of the first peak.

Peak delay

In the curve fitting results I find that the temporal alignment of pulse peaks shows some variation relative to its associated syllable and corresponding vowel onset. but the effect is small and so far nothing systematic about it has emerged.

Surface peaks on the other hand do behave in a fairly regular fashion. Figure 7 compares three words with stress and acute accent on the first syllable differing in the number of syllables per word: 4 , 40 , 400 as in '*sup*', '*fjäd*er' och '*kritiker'*. As more syllables is appended to the first stressed one the observed peak is progressively delayed. The second peak of words with the grave accent under goes a similar shift, e.g. '*anmäl'*. '*anmäla''*, '*anmälare'''*. This phenomenon can be given an explanation in terms of the pulse analysis.

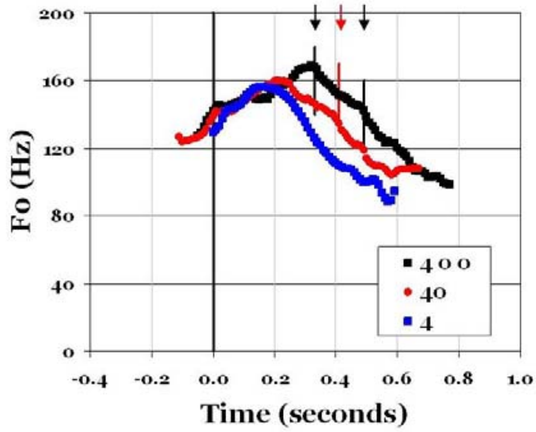

Figure 8. F0 peak delay as a function of the number of syllables following. Zero time occurs at the vowel onset in the first syllable. The vertical arrows indicate later vowel onsets..

Figure 8 reveals that, as the pulses of the following unstressed syllable(s) are added to the falling slope of the stressed pulse, the surface peak is delayed relative to its underlying pulse peak,

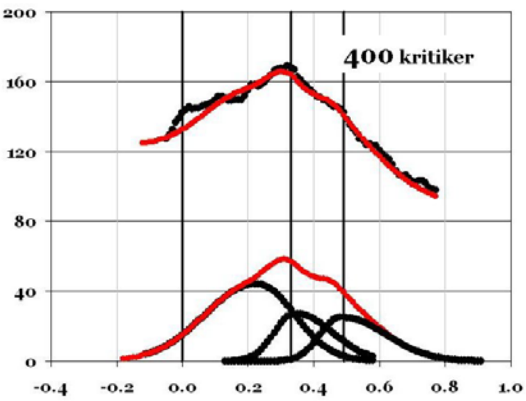

Figure 9. Pulse analysis of a '400' pattern, 'kritiker' demonstrating how adding unstressed pulses on the falling slope of the stressed pulse delays the surface peak relative to its underlying pulse peak.

Discussion

Independent motivations for pulses

Fitting the pulse model to the present F0 data was easy and could done with reasonable accuracy. It may be remarked that this is not too surprising given the number of free parameters available. Still we need to ask if the model may not also have a more interesting, independent motivation. Perhaps it works so well because it is compatible with general facts about the nature of stress and syllables?

Nature of stress

Let us begin even further back than Stetson looking at Jespersen's view of stress. He writes (1926:119): *"Akzent (Druck) ist Energie, intensive Muskeltätigkeit, die nicht an ein einzelnes Organ gebunden ist, sondern der gesamten Artikulation ihr Gepräge gibt."* In other words he views *"Druck als Gesamtenergie"* (p 120). Let me elaborate on this a bit.

It appears clear that the motor organization of a syllable involves simultaneously coordinating articulatory, phonatory and respiratory processes. Activating this three-part system inevitably requires physiological effort implying that also unstressed syllables have some degree of 'stress'. It is also evident that the stress parameter is not a continuous function of time but is discretely updated once every syllable.

According to Jespersen greater stress radiates more energy into all subsystems. As a result articulatory movements become more extensive, vocal fold tension increases and traces of a higher level of respiratory performance can be detected in EMG and P_s records. This scenario is summarized in Figure 10.

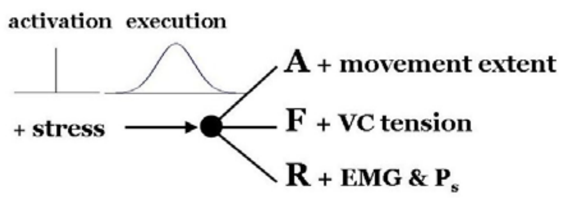

Figure 10. An attempt to summarize Jespersen's theory of stress and to identify a possible origin of the F0 pulses postulated in the present study.

Speculations

Where does the F0 pulse come from?

A possible hypothesis: From the stress pulse which controls the activity level not only of the respiratory system but also of phonatory and articulatory processes. I claim that the rise-fall charateristics of the F0 pulse derive from a stress-induced increase in vocal fold tension (and a slightly higher P_s).

Where did the tonal accents come from?

A second hypothesis: From the stress (and F0) pulses (assumed to be present in Old Norse) which in main stress syllables split into two variants: one syllabically early (grave) pulse and one syllabically late (acute) pulse The distinction may have been in place soon after the first millenium (Riad 2005, 2006).

Postscript

The present project owes a great deal to Öhman's work on word and sentence intonation (1967). His is the first attempt to model F0 contours quantitatively using step function control commands and smoothing filters. It also proposed physiological interpretations and discussed the model's implications for historical and dialectal developments of Scandinavian tone accents. A truly pioneering effort.

References

Abrahamsson M & Sundberg J (2007): "Subglottal pressure variation in actors'stage speech", in Rees M (ed): *Voice and Gender*, 343-47.

Bruce G (1977): *Swedish word accents in sentence perspective* Travaux de l'institut de linguistique de Lund 12. Lund: CWK Gleerup.

Elert C-C (1970): *Ljud och ord i svenskan*, Stockholm:Almqvist&Wiksell.

Jespersen O (1926): *Lehrbuch der Phonetik*, Teubner:Leipzig.

Ladefoged (1967): *Three Areas of Experimental Phonetics*, London:Oxford University Press.

Ladefoged P & Loeb G (2002): "Preliminary experiments on respiratory activity in speech", http://www.linguistics.ucla.edu/people/ladefoge/LadefogedAndLoebRespiration.pdf

Ladefoged P (2005): "Speculations on the control of speech", 3-21 in Hardcastle W J & Mackenzie Beck (eds): *A figure of speech*, LEA : New Jersey.

Lehiste I (1970): *Suprsegmentals*, Cambridge:MIT Press.

Riad T (1998). "The origin of Scandinavian tone accents", *Diachronica XV:1*, 63-98.

Riad T (2005). "Historien om tonaccenten", 1-27 in Falk C & Delsing L-O (eds): *Studier i svensk språkhistoria 8*, Lund: Studentlitteratur.

Riad T (2006): "Scandinavian accent typology", *Sprachtypol Univ Forsch (STUF)*, Berlin 59:36-55.

SAOB = *Svenska Akademiens Ordbok*, Lund 1898.

Stetson R H (1951): *Motor Phonetics*, Amsterdam:North-Holland.

Sundberg J, Iwarsson J & Hagegård H (1995): "A singer's expression of emotions in sung performance",. 217–229 in Fujimura O & Hirano M (eds): *Vocal Fold Physiology: Voice Quality Control*, Singular, San Diego.

Öhman S (1967): " Word and sentence intonation: A quantitative model".*STL/QPSR 8(2-3):20-54.*