A pilot study on the F0 curve of syllable-initial sonorants, comparing nasals, lenis stops and fortis stops

Alexis Michaud, Barbara Kühnert

A pilot study on the F_0 curve of syllable-initial sonorants, comparing nasals, *lenis* stops and *fortis* stops

Alexis Michaud & Barbara Kühnert

Laboratoire de Phonétique et Phonologie, UMR 7018 CNRS/Sorbonne Nouvelle & Institut du Monde Anglophone, Université Paris 3

Abstract

Several publications raise the issue whether the F_0 curve of syllable-initial sonorants can play a prosodic role. The experimental evidence adduced in the present pilot study consists of 15 C_1VC_2 words, where C_1 = /p/, /t/, /k/, /b/, /d/, /g/, /m/ or /n/, V = /ɑː/, /iː/, /uː/, and C_2 = /t/; these words were said twice inside a carrier sentence by four Cambridge undergraduates (speakers of Southern Standard British English). Comparison of the F_0 curves of the /m/-initial syllables with those of the obstruent-initial syllables suggests that only the part of the F_0 curve which corresponds to the syllable rime is to be taken into account at the stage of the interpretation of the word's intonation pattern.

1. INTRODUCTION

1.0. Some preliminary definitions

Prosody as defined here consists of *accentuation, intonation,* and several *performance factors* (such as speaking rate). *Accentuation* includes all nonphonemic lexically distinctive properties, e.g. stress in English, tone in Mandarin, or pitch accent ("tonal stress") in Japanese and Swedish. *Intonation*, which is often (and perhaps somewhat abusively) identified with the parameters whereby it manifests itself—and especially with fundamental frequency—, is a complex, abstract structure, which essentially reflects syntax (in the broader sense) and information structure; it also comprises *attitudinal* and *emotional* dimensions, that convey speaker attitudes and emotions. Further detail is proposed in Vaissière and Michaud 2006; this approach is (in our view) close to those of Coustenoble and Armstrong 1937, Delattre 1966, Rossi 1999 and Niebuhr 2013.

1.1. The prosodic role of the F_0 curve of syllable-initial sonorants

It is generally accepted that fundamental frequency during voiced stops is not controlled by the speaker. Maintaining voicing during stop consonants is difficult due to the buildup of air pressure behind the locus of constriction; *a fortiori*, it appears very unlikely that fundamental frequency (the rate of vibration of the vocal folds, hereafter F_0) can be controlled during these consonants. On the other hand, it is possible to realise F_0 modulations during the articulation of a sonorant consonant such as [m] (witness onomatopoeia such as “Mmh!”). Then again, this does not entail that the portion of the F_0 curve that is carried by a sonorant consonant necessarily plays a linguistic role. ‘Linguistic role’ is here understood very broadly, to refer to a role at the *lexical* level, i.e. participating in the realisation of lexical tone or lexical stress, or at the intonational, *supralexical* level.

This issue has a bearing on prosodic studies, insofar as a *nasal+vowel* syllable in the course of which F_0 increases during the nasal onset and decreases during the vowel can, depending on whether the portion of the F_0 curve corresponding to the initial consonant is taken into account or not, be described as having a rising-falling F_0 pattern, or simply a falling one. This is illustrated by figure 1, taken from the Aix-MARSEC corpus (Auran et al. 2004), where the F_0 increases during the /m/ of *emerged* and decreases during the following vowel.

The perspective proposed here is inspired by some reflections on the lexical tones of some East/Southeast Asian languages (set out in section 1.2). Like these lexical tones, English stress is lexically attached to the syllable; the questions raised by the domain of realisation of the two sets of phenomena appear similar to a certain extent.

1.2. The situation in prosodic systems that possess lexical tones: the syllable is divided into onset and rime; the tone-bearing unit is the rime

In the study of the phonologically monosyllabic languages of East/Southeast Asia, which have a relatively simple syllabic structure, the syllable is traditionally divided into *onset* and *rime*. In a syllable such as /man/, the onset is /m/ and the rime is /an/, i.e. syllable-final sonorants are part of the tone-bearing unit whereas syllable-initial consonants (including sonorants) are not part of it. This situation is epitomised by Vietnamese (Tran 2007:48-50; Brunelle 2009:83).

This division is verified in numerous languages (with some complexities introduced by *medial* consonants, such as the semi-vowel in syllables such as /mjan/; for more details on the traditional analysis of Chinese syllables, see Sagart 1993:35 and references therein). The tone is lexically associated to the syllable, but is carried by the rime. The prosodic description of many languages echo this division into onset and rime (for Yoruba, a language of the Niger-Congo family, see Laniran 1992:61; for Danish: Gårding 1998:137; for the Yi/Ni languages—Sino-Tibetan family—: Burling 1967:56; on the notion of *tone-bearing unit*: Clements and Ford 1979:181).

1.3. The syllable-alignment approach: taking the F_0 curve of the entire syllable into account

Several studies, which bear on a broad range of languages, consider the F_0 curve carried by sonoral initial consonants to play a part in the interpretation of the syllable's intonation contour (e.g. Ladd et al. 1999 for English, Xu Yi 1998 and Xu Yi and Wang 2001 for Mandarin Chinese, Arvaniti et al. 1998 for Greek, Prieto et al. 1995 for Mexican Spanish, Prieto 2009 for Catalan, Chahal and Hellmuth 2014:391 for Arabic). The idea is that prosodic targets – defined as including lexical tones as well as intonational phenomena – are specified with reference to syllable boundaries. Xu and Liu (2006) propose "a general model of temporal organization, in which the syllable is the basic time structure that specifies the alignment of consonants, vowels, tones (...)". Likewise, a study of child speech in English, Spanish and Catalan only uses the onset and end of accented syllables as segmental landmarks, rather than the boundary between the initial consonant and the vowel part of the syllable, i.e. between initial and rhyme (Astruc et al. 2013). These studies rest on the idea that, at a certain level of abstraction, F_0 is continuous; if it does not appear as continuous in the acoustic signal, that is due to the absence of voicing of some of the segments, not to the absence of an *underlying* F_0 curve.

On this basis, voiced sonorant consonants are preferred in the experiments set up by these authors. Several experiments focus on the 'segmental anchoring of F_0 curves', i.e. they study fine details in the temporal alignment of F_0 curves with the 'segments' (vowels and consonants). This is exemplified by the study of the semantic nuances associated with various alignments of an F_0 peak with the syllable /lo/ in the German sentence *Sie hat ja gelogen*, which concluded that "Early F_0 peaks signal established facts, middle peaks new information, and late peaks emphasis and contrast" (Kohler 1991:307; this study was followed and complemented by other studies, such as Niebuhr 2003).

This view of things contradicts the conception set out in section 1.2, without actually entering into an open debate with it. For instance, a study puts forward the idea that, in Mandarin Chinese, the F_0 curve and the segments are aligned by the beginning of the syllable, i.e. that the tone-bearing unit is the entire syllable (Xu Yi and Wang 2001:321), a suggestion which contradicts the findings of several experimental studies of the same language (in particular Howie 1974 and Hallé 1994).

The pilot study presented here is based on a small body of experimental data on "Received Pronunciation"/"Standard Southern British English". It consists in comparing the F_0 curve of two sets of monosyllables, the ones with sonorant onsets, the others with stop onsets. The results will be discussed in light of a comparison with the phonetic realisation of lexical tone in some Far Eastern languages.

2. METHOD

2.1. Corpus and speakers

Four speakers (hereafter M1 to M4) were recorded at the Phonetics Laboratory of the University of Cambridge. They were undergraduate students at the Department of Linguistics. They were retributed for their participation. We focus here on three series of consonants: *fortis* /p/, /t/, /k/ (realised as **aspirated** when in onset position, in a stressed syllable); *lenis* /b/, /d/, /g/ (realised as **unvoiced** in the same context); and the nasals /m/ and /n/.[1] In an effort to give a reasonably large scope to the

[1] In a previous study (reported in Michaud and Kühnert 2006), only labials /p/, /b/ and /m/ were studied, for the sake of the homogeneity of the results; for the

experimental paradigm, disyllables as well as monosyllables were used, so that the results could be checked across these two word structures. The 48 items recorded are

- twenty-four monosyllabic (C_1VC_2) lexical words, with C_1 being one of /p, t, k, b, d, g, m, n/, V one of /ɑː/, /iː/, /uː/, and C_2 a final stop: e.g. *boot, cart, deep*;
- twenty-four disyllabic ('C_1VC_2ə) lexical words, of the same phonemic composition as before, but with an added /ə/: e.g. *booter, Carter, deeper*.

The words were read twice inside the same carrier sentence, with different indications of context (chosen for the sake of a cross-language comparison; see Michaud 2005:107-128, 381-390). The first task will hereafter be referred to as condition C (for Careful) and the second one as condition E (for Emphatic; needless to say, these labels are chosen simply for convenience). These two different attitudinal conditions are a source of variation—a variation above which the linguistic regularities under investigation should emerge.

Task 1: You're teaching a foreign student who made a mistake when reading a word. (Class context.) Read each item inside the carrier sentence, making a long pause (breathing in and out once) in-between sentences: *Look, this is ___ here.*

Task 2: A child who is learning to read has asked you how to pronounce this word time and again; (s)he asks you yet another time; you answer, less patiently: *Look, this is ___ here!* Remember to make a long pause (breathing in and out once) in-between sentences.

Both sentences being statements, the expected pattern within the traditional British notation of intonation would be a *Fall*.

In order to obtain a very precise measurement of F_0 (as well as indications on voice quality, not used here), an electroglottographic recording was conducted simultaneously with the audio recording.

2.2. Data analysis

The beginning and end of each syllable and the boundaries between initial nasal consonants and vowels were determined on the basis of inspection of the audio signal, detecting the abrupt change in the shape of the signal at the transition between the nasal onset and the vowel (see figure 2 for an illustration). Segmentation was easy, as the reading

---

purpose of the present study, labials, coronals and velars are bunched up together in the results.

conditions elicited an especially clear *(hyper-articulated)* rendering. In dubious cases, segmentation was verified spectrographically. (The software used were SOUNDFORGE, www.sonicfoundry.com, and PRAAT, www.praat.org.) F_0 was calculated by detecting the positive peaks ("glottis-closure peaks") on the derivative of the electroglottographic signal, under the computing environment MATLAB (www.mathworks.com). On this method, see Henrich et al. 2004 and references therein; the MATLAB routines implementing the method, developed by the first author of this paper (and described in Michaud 2004), are available for download from the GitHub repository COVAREP (Degottex n.d.). Requests for the MATLAB routine created for calculating average curves for each data subset should be sent to the authors.

The present pilot study rests solely on observations on the time course of F_0; this is a limitation, given the perceptual role of other parameters, such as loudness and voice quality. A refined understanding of the present issue would require a multiparametric investigation, as well as a perceptual evaluation.

3. RESULTS AND DISCUSSION

3.1. Comments on the figures

Figure 3a shows, for one speaker (M1) and one reading condition (condition C: careful reading), the F_0 curves of the three types of monosyllables (with fortis stops, lenis stops, and nasals). Each point corresponds to one glottal cycle. Figure 3b shows the same data without indicating standard deviation, in order to make the figure more legible. Figure 4 shows the results for condition E (emphatic reading). Figures 5 and 6 represent data from another speaker. In order to evaluate the degree of proximity between the curves obtained (i) on nasal-initial syllables when the initial nasal is disregarded in the calculation, (ii) on syllables beginning by a *lenis* stop, and (iii) on syllables beginning by a *fortis* stop, these three data subsets were aligned by their point of origin.

Looking at the F0 patterns on the rimes in figures 3-6, the following differences emerge: the fall realised on syllables with initial fortis consonants generally begins from a higher value than after a lenis or nasal consonant, and tends to have a stronger downward slope. This may be put down to the well-attested aerodynamic and physiological effects of initial aspiration, an effect often referred to as 'micromelodic' or 'microprosodic' (see, e.g., Di Cristo 1985).

The curves calculated over the whole length of the nasal-initial syllables (thin line) consistently show a gradual rise (or a flat portion of curve) that precedes the decrease of F_0. This flat or rising portion of curve is considerably longer than that occasionally observed on syllables with onset /b, d, g/ (squares).

For all conditions, final glottalisation occurs in some cases (see figure 3), resulting in irregular glottal cycles; this can be overlooked for the purpose of the present investigation.

Overall, the F_0 curves calculated over the rimes exhibit a stronger resemblance than those calculated over entire syllables.

3.2. Similarity of the F_0 curves over the rimes

The context of utterance being the same for all the items, their F_0 curves correspond, at a certain level of abstraction, to the same linguistic phenomenon: a certain kind of realisation of word stress. This phenomenon would be categorised in different ways in different frameworks: for instance, as a Fall in the British tradition of intonational studies, or as a H*L tone sequence in a ToBI-style notation (ToBI, for Tones and Break Indices, is a system of notation of intonation which was proposed as a standard by Silverman et al. 1992, and which is still very influential, despite major shortcomings pointed out by Wightman 2002 and Vaissière 2002, among others). As the choice of one model or another does not have a direct bearing on the present issue, the phenomenon under investigation will simply be referred to, in the abstract, as ‘A’. The differences between the curves of the stop-initial syllables and those of the nasal-initial syllables can be interpreted in at least two ways.

(i) Under the first view, the curve of sonorant-initial syllables provides the most complete image of A, whereas the F_0 curve of syllables with an unvoiced onset is incomplete: an F_0 pattern is specified for the syllable as a whole, and aligned with the beginning and end of the syllable; part of it is not phonetically realised because of the consonant’s features (obstruent, and unvoiced).

(ii) The curve of obstruent-initial syllables provides a complete image of A; as for the curve of nasal-initial syllables, only that portion of the curve which corresponds to the rime needs to be retained in order to bring out the portion of the F_0 rime which corresponds to A. The flat or rising F_0 curve during the initial consonant reflects a transition in-between A and the previous prosodic events (in the present case, the rise is to be interpreted as an anticipatory movement towards the starting-point of a falling curve).

Following interpretation (i), which we believe corresponds to the orientation set out in section 1.3, one is led to describe A as made up of two parts: a stable part (slightly rising) followed by a fall, resulting in what could be considered as a Rise-Fall. This conclusion does not accord with the British framework of intonational studies (e.g. Palmer 1922, Kingdon 1958, O'Connor and Arnold 1973), in which the realisation expected in this context is a Fall. On the other hand, going by interpretation (ii), which is inspired by the traditional division of the syllable into onset and rime, A is a Fall in all cases: as pointed out previously, once the F_0 curve of nasal-initial syllables is simplified to the part which corresponds to the rime, a greater degree of similarity (in terms of length and of shape) is observed between the three subsets (with nasal onset, with *fortis* stop onset, and with *lenis* stop onset). The considerable difference in length between the curves with initial nasal and those with initial stop by and large corresponds to the length of the initial nasal consonant. These visual observations are in keeping with the traditional notion that the rime, not the entire syllable, is the carrier of prosodic phenomena (lexical tones, stress, and intonational phenomena).

3.3. Hypotheses on perception

Recent research has established that listeners can make use of tenuous phonetic clues to lexical oppositions (see, in particular, Hawkins 2003); this might cast doubt on the validity of a stylisation that excludes the portion of the F_0 curve carried by an initial consonant. But in fact, this choice does by no means amount to denying that syllable-initial consonants can play an intonational role, by their degree of length, and their articulation (at the supraglottal, glottal and subglottal levels), as shown by Fónagy 1983, among others, who investigates the importance of consonantal articulation in the expression of emotions. Moreover, it appears plausible that listeners can use the F_0 curve of syllable-initial sonorants as a secondary cue to the syllable’s melody: e.g. a rise during a syllable-initial /m/ could help perceive a fall on the following vowel.

It must also be acknowledged that the present conclusion is based only on a very limited spectrum of intonational functions. Looking at other intonational categories, Gartenberg and Panzlaff-Reuter 1991 and van Santen and Hirschberg 1994 observed quite different F_0 courses for accented syllables with and without a voiced initial, beyond the differences expected on the basis of the intrinsic characteristics of these consonants. These findings show that simply truncating F_0 curves in the way it was performed for the sake of our demonstration cannot be recommended in every case as an adequate way of factoring out the effect of initial consonants.

4. CONCLUSION

The results of the experimental pilot study lead us to conclude that, in the case of the English monosyllables under study, leaving aside the portion of the F_0 curve which is carried by the initial consonant /m/ makes the prosodic phenomenon realised on the syllable emerge more clearly than when the curve for the entire syllable is taken into account.

Our observations evidently do not mean that F_0 movements during initial sonorants, and F_0 perturbations due to initial consonants, are not used by listeners as cues; but ironically, this information, whilst used for gaining segmental information, seems to be discarded at the level of the perception of prosody (Reinholt Petersen 1986).

5. ACKNOWLEDGMENTS

Many thanks to Francis Nolan for allowing use of the facilities of the phonetics lab at Cambridge University, to Geoffrey Potter for technical help, and to the speakers for participating in the experiment. Many thanks to Oliver Niebuhr, and to our anonymous reviewers, for detailed comments and helpful suggestions.

References

Arvaniti, A.; Ladd, R.; Mennen, I.: Stability of tonal alignment: the case of Greek prenuclear accents. Journal of Phonetics **26**: 3-25 (1998).

Astruc, L.; Payne, E.; Post, B.; Vanrell, M.; Prieto, P.: Tonal targets in early child English, Spanish, and Catalan. Language and Speech **56**(2): 229–253 (2013).

Auran, C.; Bouzon, C. & Hirst, D. (2004). "The Aix-MARSEC Project: An Evolutive Database of Spoken British English." Proceedings of the Second International Conference on Speech Prosody, Nara, Japan, pp. 561-564.

Brunelle, M.: Tone perception in Northern and Southern Vietnamese. Journal of Phonetics **37**: 79-96 (2009).

Burling, R.: Proto-Lolo-Burmese (Mouton & Co., Bloomington/The Hague 1967).

Chahal, D.; Hellmuth, S.: The intonation of Lebanese and Egyptian Arabic. Prosodic typology II: the phonology of intonation and phrasing: 365–404. S.-A. Jun. (Oxford University Press, Oxford 2014).

Clements, N.; Ford, K.C.: Kikuyu tone shift and its synchronic consequences. Linguistic Inquiry **10**: 179-210 (1979).

Coustenoble, H.; Armstrong, L.: Studies in French intonation (Heffer, Cambridge 1937).

G. Degottex: COVAREP: A Cooperative Voice Analysis Repository for Speech Technologies (n.d.). http://covarep.github.io/covarep/

Delattre, P.: Les dix intonations de base du français. The French Review **40**(1): 1-14 (1966).

Fónagy, I.: La vive voix: essais de psycho-phonétique (Payot, Paris 1983).

Gårding, E.: Intonation in Swedish. Intonation Systems: A Survey of Twenty Languages: 112-130. D. Hirst and A. Di Cristo. (Cambridge University Press, Cambridge 1998).

Hallé, P.: Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese. Language and Speech **37**: 103-124 (1994).

Hawkins, S.: Roles and representations of systematic fine phonetic detail in speech understanding. Journal of Phonetics **31**: 373-405 (2003).

Henrich, N.; d'Alessandro, C.; Castellengo, M.; Doval, B.: On the use of the derivative of electroglottographic signals for characterization of non-pathological voice phonation. Journal of the Acoustical Society of America **115**(3): 1321-1332 (2004).

Howie, J.M.: On the domain of tone in Mandarin. Phonetica **30**: 129-148 (1974).

Kingdon, R.: The Groundwork of English Intonation (Longman, London 1958).

Kohler, K.J.: Terminal intonation patterns in single accent utterances of German: phonetics, phonology, and semantics. Studies in German Intonation: 115-185. (University of Kiel (Germany) 1991).

Ladd, R.; Faulkner, D.; Faulkner, H.; Schepman, A.: Constant "segmental anchoring" of F_0 movements under change in speech rate. Journal of the Acoustical Society of America **106**: 1543-1554 (1999).

Laniran, Y.O.: Intonation in Tone Languages: the Phonetic Implementation of Tones in Yorùbá (Ph. D., Cornell University, Ithaca, New York 1992).

Michaud, A.: Final consonants and glottalization: new perspectives from Hanoi Vietnamese. Phonetica **61**(2-3): 119-146 (2004).

Michaud, A.; Kühnert, B.: La courbe de F_0 des sonantes initiales de syllabe joue-t-elle un rôle prosodique? Etude-pilote de données d'anglais britannique. Proceedings of *XXVI^e Journées d'Etude de la Parole*, Dinard: 121-124 (2006).

Niebuhr, O.: Perceptual study of timing variables in F_0 peaks. 15th International Congress of Phonetic Sciences, Barcelona: 1225-2118 (2003).

Niebuhr, O.: The acoustic complexity of intonation. Nordic Prosody XI: 15–29. E.L. Asu and P. Lippus. (Peter Lang, Frankfurt 2013).

O'Connor, J.D.; Arnold, G.F.: Intonation of Colloquial English (Longman, London 1973).

Palmer, H.E.: English Intonation, with Systematic Exercises (Heffer, Cambridge 1922).

Prieto, P.:. Tonal alignment patterns in Catalan nuclear falls. Lingua **119**(6): 865–880 (2009).

Prieto, P.; van Santen, J.; Hirschberg, J.: Tonal alignment patterns in Spanish. Journal of Phonetics **23**: 429-451 (1995).

Rossi, M.: L'intonation, le système du français: description et modélisation (Ophrys, Gap/Paris 1999).

Sagart, L.: Les dialectes gan. Etudes sur la phonologie et le lexique d'un groupe de dialectes chinois (Langages croisés, Paris 1993).

Silverman, K.E.A.; Beckman, M.; Pitrelli, J.; Ostendorf, M.; Wightman, C.W.; Price, P.; Pierrehumbert, J.; Hirschberg, J.: ToBI: A Standard for Labeling English Prosody. Proceedings of the 1992 International Conference on Spoken Language Processing, Banff, Canada: 867-870 (1992).

Tran, D.-D.: Tran, Do-Dat. 2007: Synthèse de la parole à partir du texte en langue vietnamienne. Ph.D. Laboratoire d'Informatique de Grenoble and Hanoi University of Science and Technology (2007).

Vaissière, J.: Cross-linguistic prosodic transcription: French vs. English. Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. L.V. Bondarko: 147-164. N. B. Volskaya, N. D. Svetozarova and P. A. Skrelin. (St Petersburg State University Press, St Petersburg 2002).

Vaissière, J.; Michaud, A.: Prosodic constituents in French: a data-driven approach. Prosody and syntax: 47-64. I. Fónagy, Y. Kawaguchi and T. Moriguchi. (John Benjamins, Amsterdam 2006).

Wightman, C.W.: ToBI or not ToBI? Proceedings of Speech Prosody 2002, Aix en Provence (2002).

Xu, Y.: Consistency of tone-syllable alignment across different syllable structures and speaking rates. Phonetica **55**: 179-203 (1998).

Xu, Y.; Liu, F.: Tonal alignment, syllable structure and coarticulation: toward an integrated model. Italian Journal of Linguistics **18**: 125–159 (2006).

Xu Yi; Wang, E.Q.: Pitch targets and their realization: Evidence from Mandarin Chinese. Speech Communication **33**: 319-337 (2001).

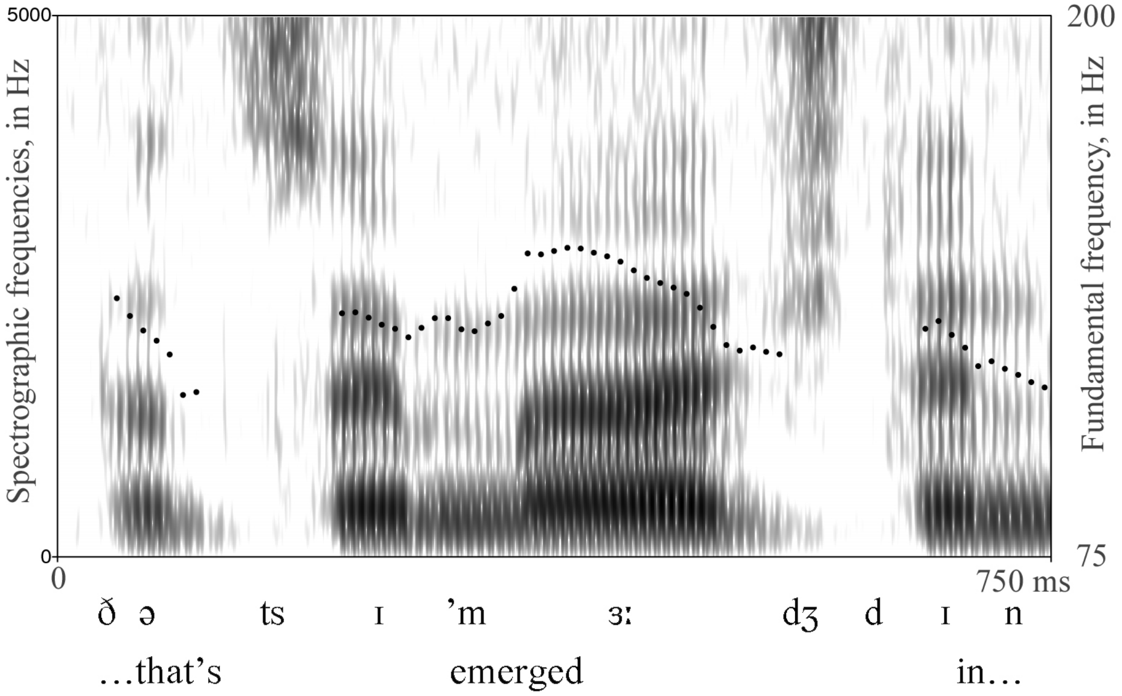

Figure 1. Spectrogram and F0 tracings of an excerpt from the Aix-MARSEC corpus, showing a rising F0 on a syllable-initial nasal followed by a decrease in F0 in the course of the syllable rhyme.

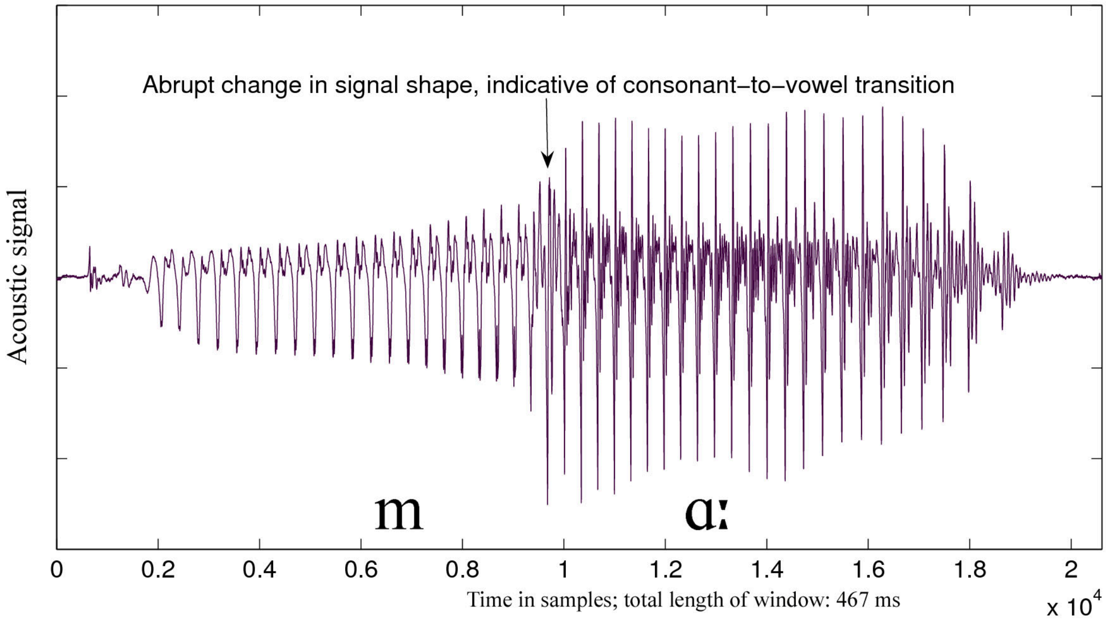

Figure 2. Illustration of the abrupt change in signal shape used as a basis for segmentation of the nasal-initial syllables. Item: *mart*, first speaker, reading condition C.

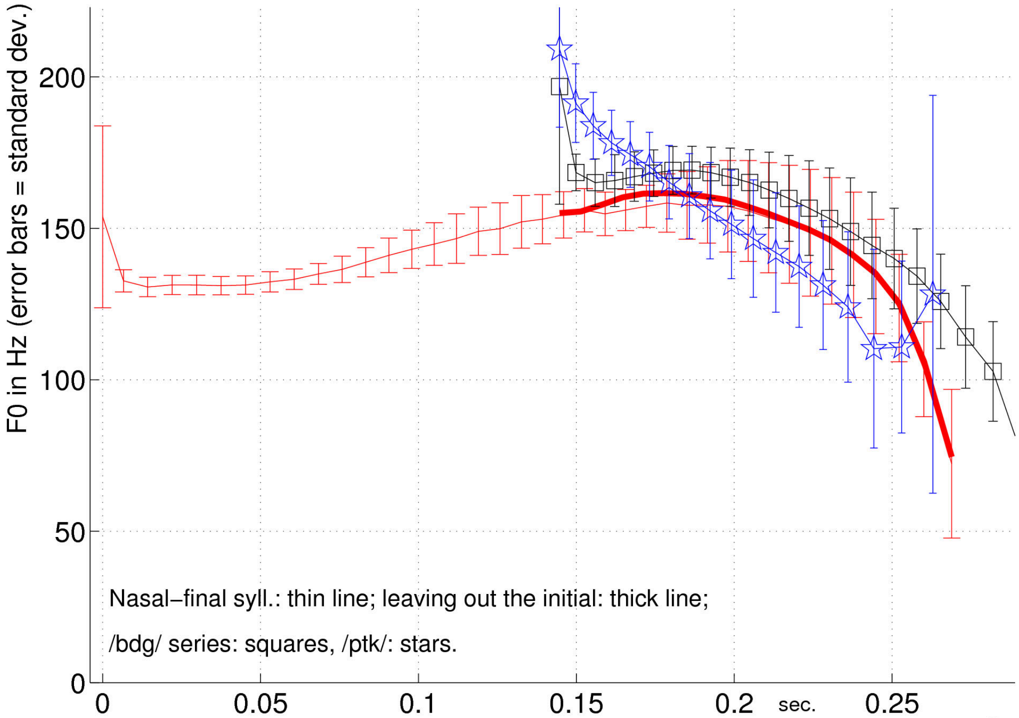

Fig. 3a. F0 curves averaged over 24 monosyllables; speaker 1, condition C.

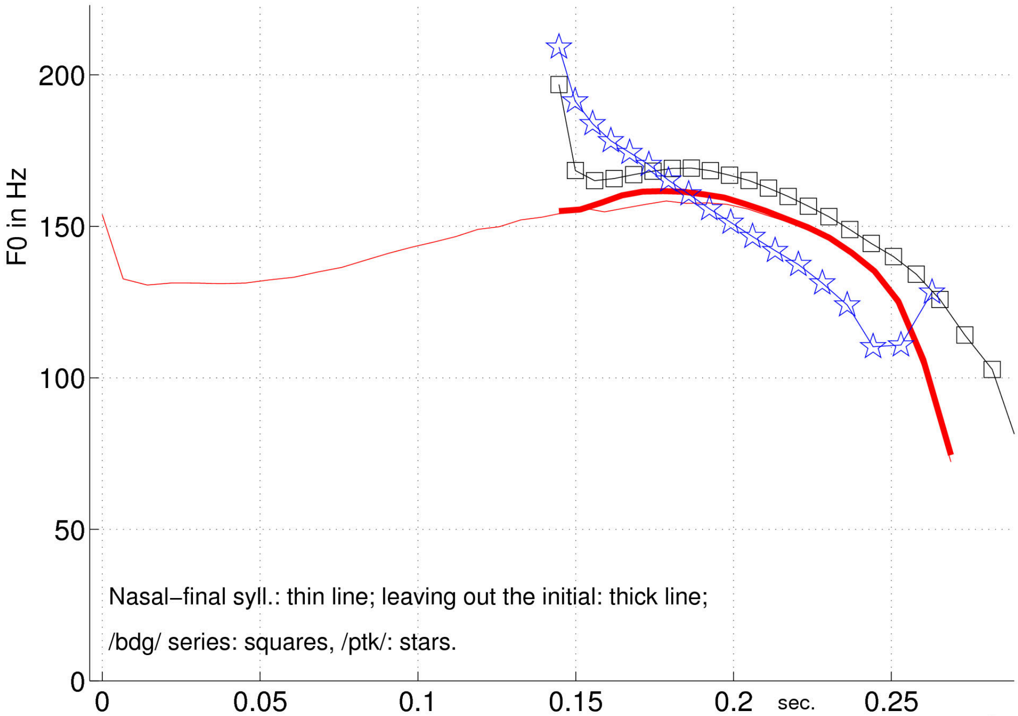

Fig. 3b. F0 curves averaged over 24 monosyllables; speaker 1, condition C.

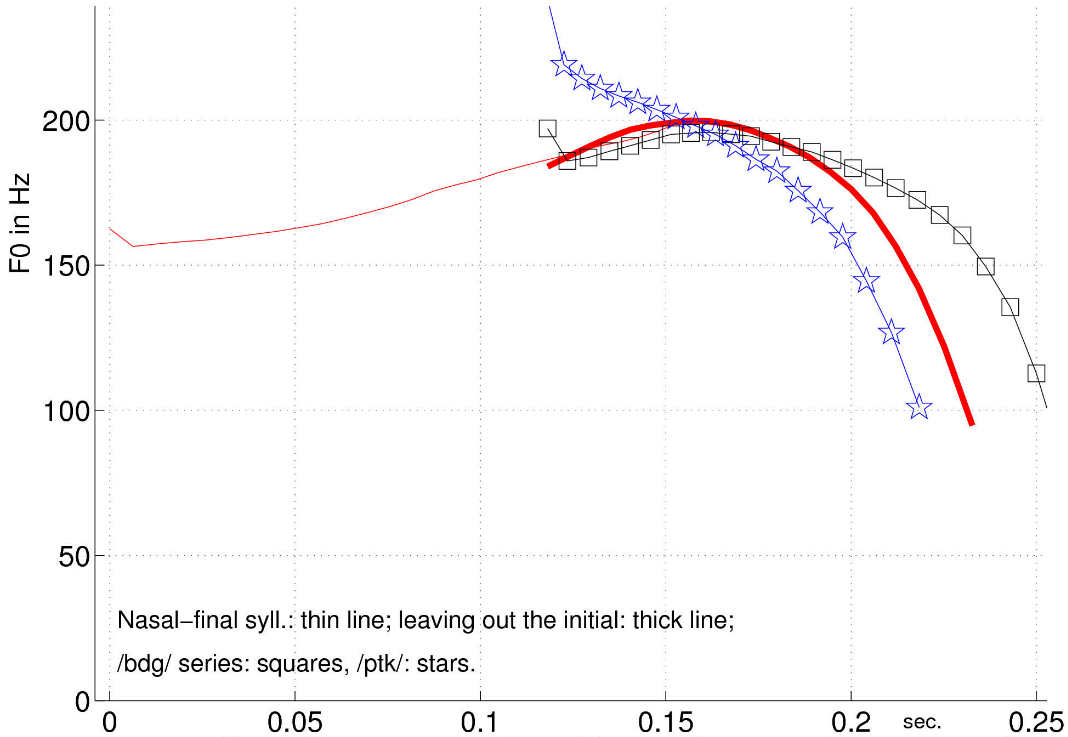

Fig. 4. F0 curves averaged over 24 monosyllables; speaker 1, condition E.

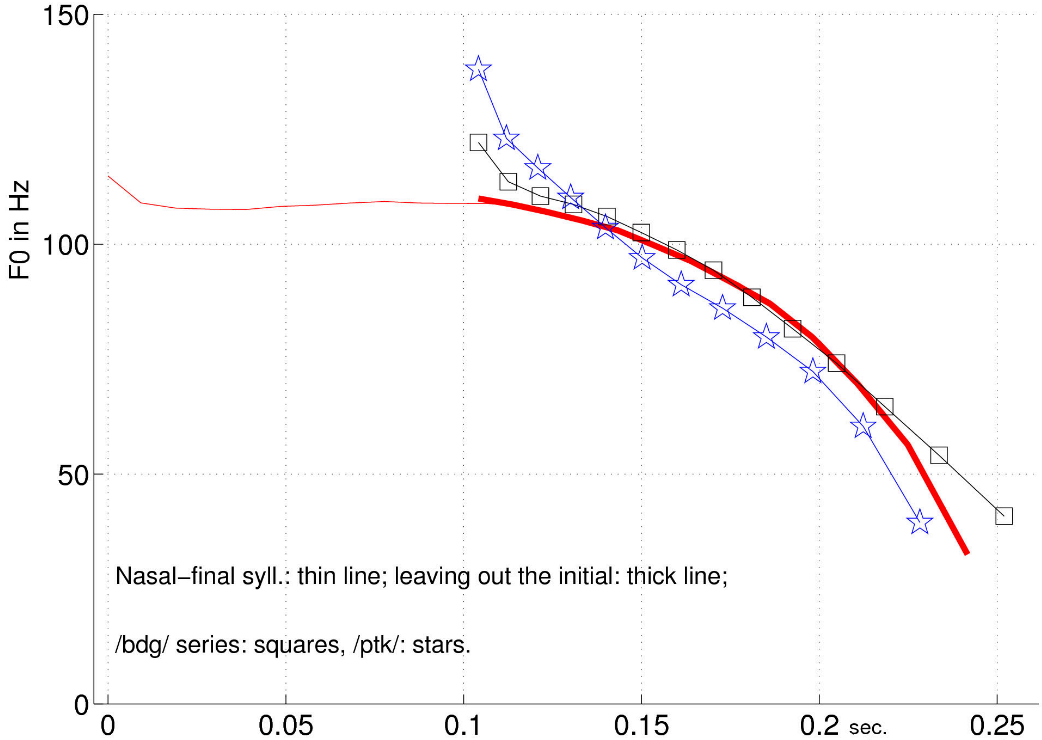

Fig. 5. F0 curves averaged over 24 monosyllables; speaker 2, condition C.

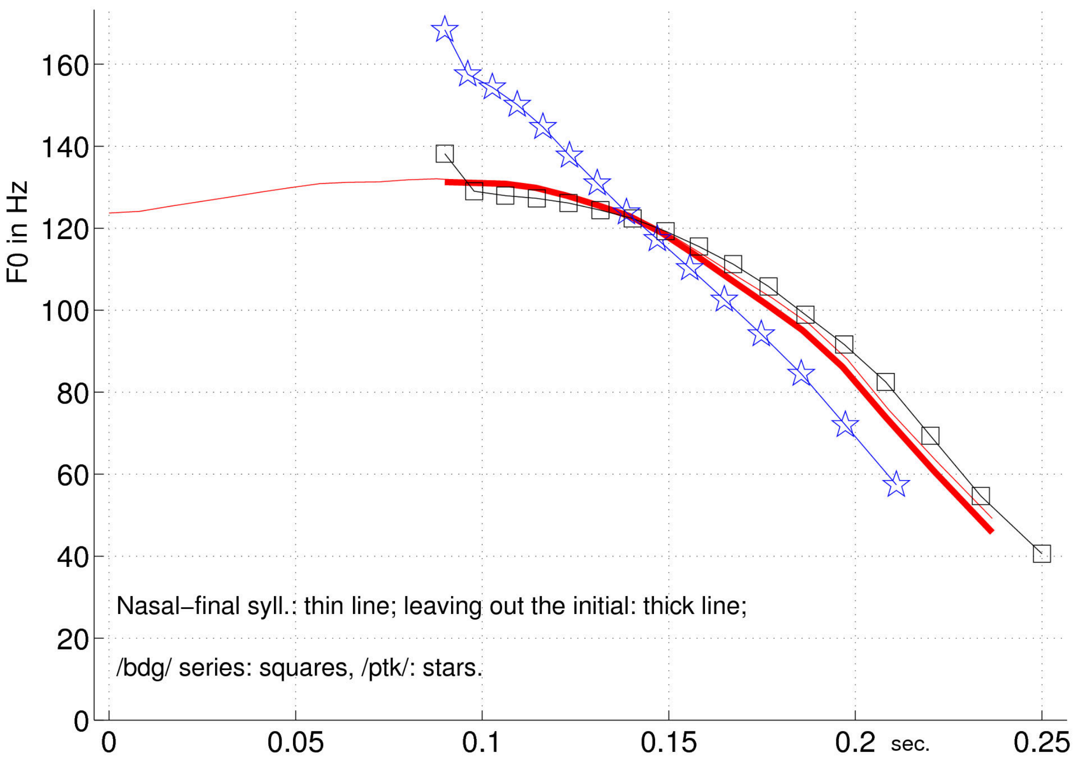

Fig. 6. F0 curves averaged over 24 monosyllables; speaker 2, condition E.