



## FormantPro as a Tool for Speech Analysis and Segmentation

### *FormantPro como uma ferramenta para a análise e segmentação da fala*

Yi Xu

Department of Speech, Hearing and Phonetic Sciences, University College London / UK  
yi.xu@ucl.ac.uk

Hong Gao

English Department, Sichuan University, Chengdu / China  
395478712@qq.com

**Abstract:** This paper introduces FormantPro, a Praat-based tool for large-scale, systematic analysis of formant movements, especially for experimental data. The program generates a rich set of output metrics, including continuous contours like time-normalized formant trajectories and formant velocity profiles suitable for direct graphical comparisons, and discrete measurements suitable for statistical analysis. It also allows users to generate mean trajectories and discrete measurements averaged across repetitions and speakers. As an illustration of its usage, data from a preliminary study of syllable segmentation in Mandarin were presented. The alignment of continuous formant trajectories enabled by FormantPro provides evidence that the temporal scopes of consonants and vowels are very different from those based on conventional views, and that acoustic and articulatory boundaries of segments are fundamentally similar.

**Keywords:** FormantPro; formant trajectories; syllable segmentation.

**Resumo:** Este artigo apresenta o FormantPro, uma ferramenta que roda no Praat, dedicada à análise sistemática e em larga escala dos movimentos de formantes, especialmente para dados de natureza experimental. O programa gera um rico conjunto de métricas de saída, incluindo contornos contínuos, como as trajetórias de formantes

normalizadas temporalmente e perfis de velocidade de formantes adequados para comparações gráficas diretas, bem como medidas discretas adequadas para a análise estatística. O programa também permite aos usuários gerar médias de trajetórias e medidas discretas calculadas a partir das médias de repetições e de falantes. Como ilustração da sua usabilidade, dados preliminares de um estudo sobre segmentação silábica em mandarim foram apresentados. O alinhamento de trajetórias contínuas de formantes geradas pelo FormantPro oferecem evidência de que os escopos temporais de consoantes e vogais são muito diferentes daqueles baseados em visões convencionais, e de que as fronteiras acústicas e articulatórias dos segmentos são fundamentalmente semelhantes.

**Palavras-chave:** FormantPro; trajetórias dos formantes; segmentação silábica

Submitted on January 8th, 2018

Accepted on June 7th, 2018

## 1. Introduction

Researchers frequently face a dilemma when it comes to taking formant measurements. Done too sparsely, important details may be missed; but continuous formant tracks are just too hard to process on a large scale. As a result, continuous formant contours are mostly used only as illustrations rather than as data in the literature. The benefit of analyzing fully continuous formants is evident from the classic work of Öhman (1966), whose insights on coarticulation gained from hand-traced formant trajectories are relevant even to the present day. But manual tracking of formants would no longer meet today's standards. Rapid technological advances have made automatically extracted continuous formants tracks easily available, yet they are still hard to use in systematic comparisons. The main difficulty is that when utterances differ in duration, it is hard to be sure whether we are comparing like with like as far as continuous trajectories are concerned.

FormantPro, available at <http://www.homepages.ucl.ac.uk/~uclyyix/FormantPro/>, is a software tool developed chiefly to address this dilemma. It is written as a Praat script (BOERSMA, 2001)—so that no programming is required of the users—for large-scale, systematic experimental studies of formant movements. FormantPro was first developed in 2007 (XU, 2007), and has been available online since 2013. The dedicated web page lists step-by-step instructions on how to use the script, how to read its output, as well as relevant information on time-normalization. The script has been used to generate results in a number of publications both by ourselves and by other researchers (e.g., BERKSON *et al.*, 2017; CHENG; XU, 2013; GAO; XU, 2013; LEE; MOK, 2017; LIU; LIANG, 2016; XU, 2007).

FormantPro applies time-normalization to extract the same number of evenly spaced formant values from each temporal interval, which allows users to treat any hypothetical unit as being temporally equivalent. The time-normalization algorithm was similar to that of ProsodyPro, a script for  $F_0$  analysis (XU, 2013), except that the default number of normalized points is 20 instead of 10 due to faster segmental than  $F_0$  changes in articulation (CHENG; XU, 2013; XU, 2007). The script further enables users to average the time-normalized formant as well as formant velocity trajectories across repetitions or even speakers. When plotted graphically, the trajectories can be compared between experimental conditions in a manner that is even more straightforward than in Öhman (1966), i.e., to overlay them in the same plot, as shown in the many examples presented later in this paper.

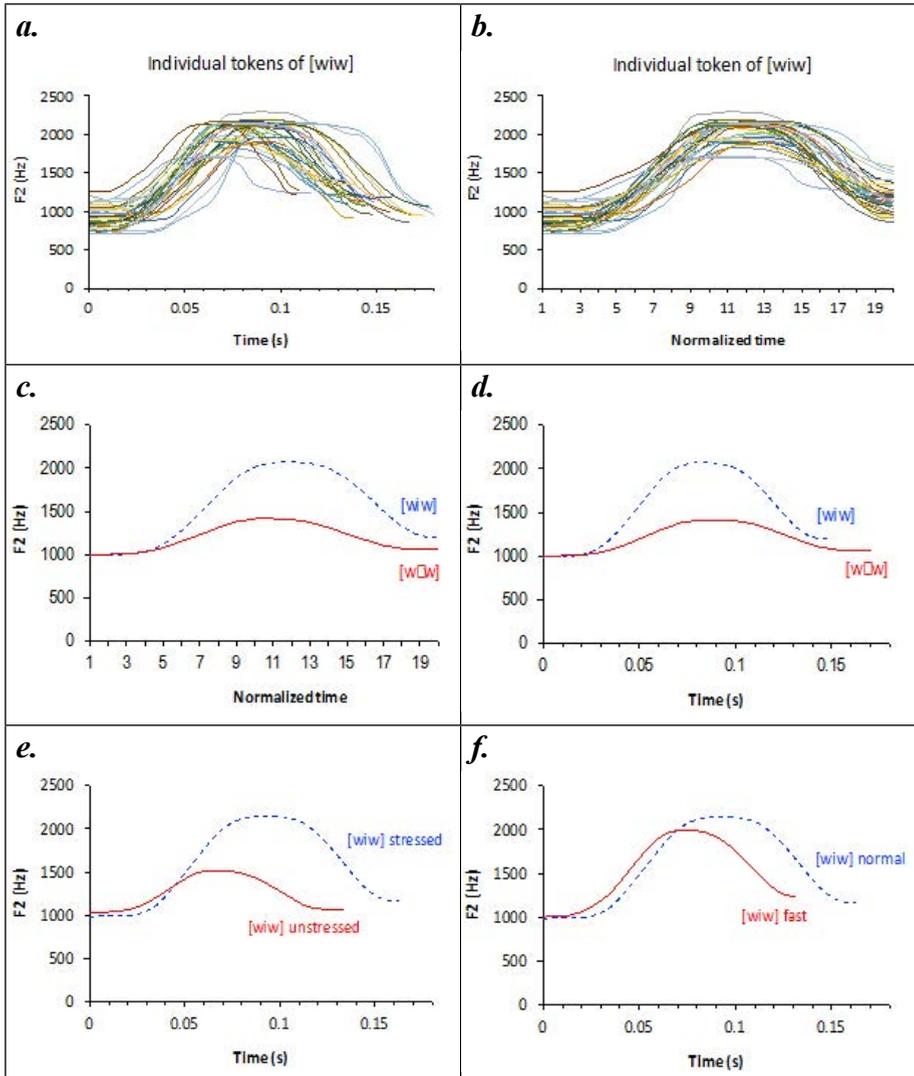
## 2. Dilemma and solution by FormantPro

Figure 1 shows F2 trajectories of [wiw] and [wow] in American English, generated by FormantPro, and plotted in three different time scales. The raw data are from utterances produced by 7 male speakers of American English. In Figure 1a, trajectories of 33 individual utterances of [wiw] are plotted in real time relative to the onset of the first F2 minimum. As can be seen, the trajectories vary extensively in duration. This makes it hard to see if there is any clear consistency across the individual tokens.

In Figure 1*b*, the same trajectories are time-normalized, i.e., consisting of 20 evenly spaced points. The consistency across the individual tokens now becomes much more apparent. The individual trajectories, however, even when time-normalized, are still not ideal for making cross-category comparisons. But given that they all consist of the same number of points, it is possible to average them at each and every point. The resulting mean trajectories can then be easily compared to each other when drawn in the same graph, as can be seen in Figure 1*c*, where the mean F2 trajectories of [wiw] and [waw] are plotted over normalized time.

A seeming disadvantage of time-normalization is that some of the original timing information may be lost. But this is not necessarily the case, as timing can also be abstracted. That is, like the formant values, the time value at each of the 20 points can be averaged across the individual tokens. In Figure 1*d*, the same [wiw] and [waw] trajectories are plotted over *averaged real time* across all individual tokens. Here the differences in terms of both curvature and timing of F2 movements between the two syllables can be easily seen. Likewise, the effects of stress and speech rate on the same syllable ([wiw]) can be easily seen when the F2 trajectories are plotted over mean averaged time in Figure 1*e* and 1*f*.

FIGURE 1 – F2 trajectories of [wiw] and [wɔw] produced by 7 male speakers of American English. In *a* and *b*, trajectories of all the individual utterances are plotted either in real time relative to the onset of the first F2 minimum (*a*) or in normalized time (*b*). In *d*, *e* and *f*, the *y* values of the trajectories are the mean F2 averaged at each of the 20 time-normalized points across all tokens by all speakers, but the *x* values are the mean times averaged also across all the repetitions and speakers at each of the 20 points. All contours are generated by FormantPro.



Another advantage of directly comparing continuous formant trajectories is that it allows one to clearly see where the largest differences are between the contrasting conditions, as the full time-course of the trajectories is immediately visible (Figure 1c-f). This enables one to make well-informed decisions when choosing measurements for statistical comparisons. Without such trajectory comparisons, decisions about where to take measurements are often made blindly, and the detection of critical differences is often a hit-and-miss game.

### 3. Usage and Features

FormantPro is written as a Praat script, which makes it executable on most of the major operating systems, including Mac, Windows and Unix. Written with large-scale systematic studies in mind, it maximizes efficiency of data processing by automating tasks that do not require human judgment, and by saving analysis output in formats that are ready for graphical and statistical analysis. More specifically, FormantPro allows users to:

- Manually segment and label intervals for each sound file, as illustrated in Figure 2,
- Cycle through all sound files in a folder without using menu commands, see Figure 3,
- Get maximum formant, minimum formant, mean formant, maximum formant velocity, duration and mean intensity from each labeled interval in each sound,
- Collect results from all individual sounds in a folder into a set of ensemble files that contain measurements of F1, F2, F3 and F2\_3 in each interval of each sound file:
  1. meanformant.txt — mean values of the formants (Hz)
  2. maxformant.txt — maximum values of the formants (Hz)
  3. minformant.txt — minimum values of the formants (Hz)
  4. maxformantvelocity.txt — maximum velocity of the formants (Hz/s)
  5. formant.txt — time-normalized formants (Hz)

- 6. `normtime_barkformant.txt` — time-normalized formants in the Bark scale (Bark)
- 7. `formantvelocity.txt` — time-normalized formant velocity (Hz/s), and
- Get mean time-normalized formants, time-normalized formant velocities and actual times corresponding to the time-normalized formant and velocity points, averaged across repetitions as well as speakers.

FIGURE 2 – A TextGrid window with hand-labelled segmentation. FormantPro generates continuous as well as discrete measurements only for the labelled intervals. This allows users to obtain measurement generation only from regions of interest.

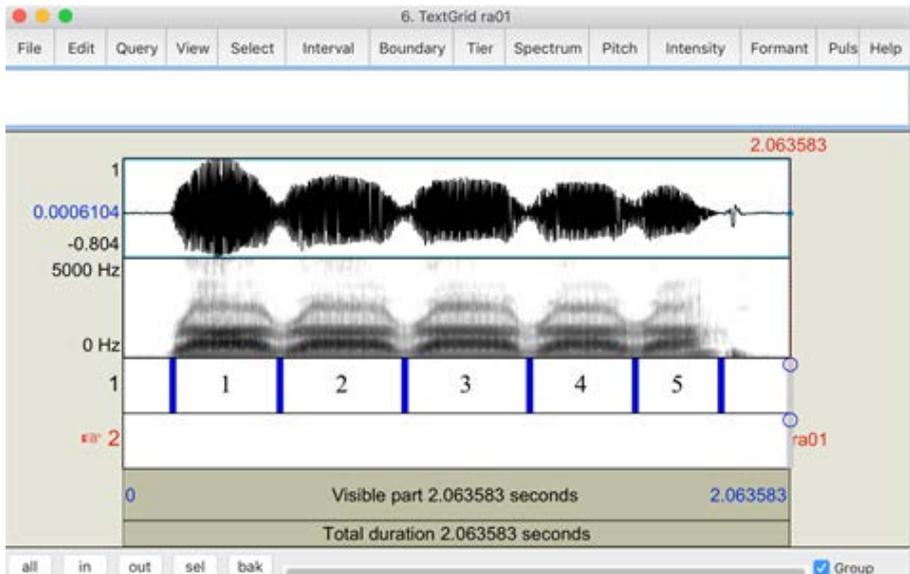
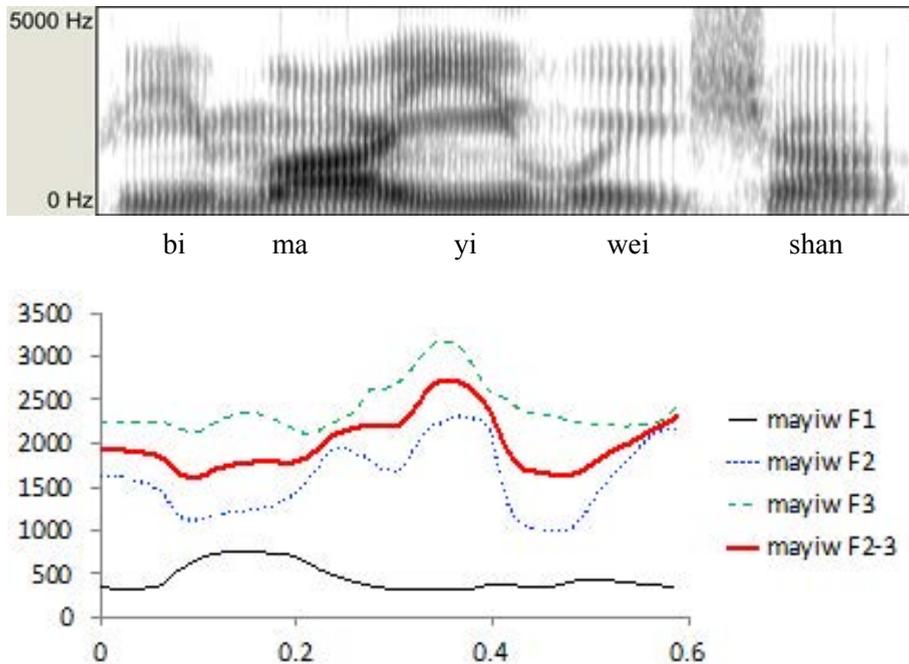


FIGURE 3 – The Pause window that controls the flow of the interactive segmentation and annotation. As the user moves forward, backward or jump to a particular sound file, the segmentation and measurements of the current file are automatically saved.



The continuous trajectories of F1, F2, F3 and F2\_3 are generated with Praat's built-in "To Formant (burg)..." function. Here F2\_3 = mean (F2, F3) is an unconventional one. It is motivated by the well-known problem of abrupt shifts of affiliation of formant with resonance cavity as vocal tract shape changes smoothly, e.g., between [i] and [a] (STEVENS, 1998). Averaging F2 and F3 can partially reduce the effects of the sudden shifts. Whether this measurement is advantageous over measuring F2 and F3 separately is an empirical matter. Data from one of our own studies (GAO; XU, 2013) seem to show partial support for this hypothesis. Making this measurement available in FormantPro will allow users to further test the hypothesis.

FIGURE 4 – Top: Spectrogram of Mandarin sentence “Bǐ Mǎyí wěishàn” [more hypocritical than Aunt Ma]. Bottom: Mean formant tracks of 10 repetitions by a male speaker of Mandarin.



Time-normalization, however, requires users to define the temporal domain of normalization. In FormantPro this is done by inserting interval boundaries in the TextGrid of an utterance. Technically FormantPro allows user to freely annotate the temporal domains of normalization, e.g., segment, syllable or even word. But meaningful time-normalization can be obtained only if there are good reasons to believe that the formant trajectories in the unit are consistently produced, which is both a theoretical and empirical matter.

To segment continuous speech into discrete units, one of the critical questions is, what is the acoustic correlate of a phonetic unit? In the current practice, the answer is that a unit, such as a consonant or vowel, is what is delimited by the landmarks (STEVENS, 2002) on a spectrogram, such as abrupt spectral shift, onset and offset of oral closure, etc. (TURK; NAKAI; SUGAHARA, 2006), which is also what sounds like that phone when isolated from the acoustic stream (ZUE *et*

al., 1990). For example, in Figure 4, the [i] in “bǐ” is to be delimited by the first and second abrupt spectral shifts after the consonant release and before the nasal murmur; the [m] in “má” is delimited by the onset and offset of the nasal murmur; and the [ʃ] is delimited by the onset and offset of the frication. This segmentation scheme, however, leaves many cases unresolved. In Figure 4, for example, the exact offset of [a], the onset as well as the offset of [ji], and the onset of [wei] is by no means clear. The vagueness of their segmentation has led to explicit advice to avoid the glides when precise duration measurements are needed (TURK; NAKAI; SUGAHARA, 2006).

From an articulatory perspective (SALTZMAN; MUNHALL, 1989; XU; WANG, 2001), however, unit boundaries can be defined rather differently. That is, the onset of a unit should be the moment when the articulators start to move toward their target positions defined by its canonical form, and the offset of the unit should be the moment when the articulators start to move away from those positions. The canonical form of a monophthong vowel would be the ideal vocal tract shape that generates the steady-state prototypical formant pattern, and the canonical form of a consonant would be the ideal closure or constriction at the appropriate place of articulation. The movements toward these targets take time, and *it is the time course of the movement that should be considered as the interval of the unit* (XU; LIU, 2007). In other words, a unit is delimited by the onset and offset of the movement toward its target.

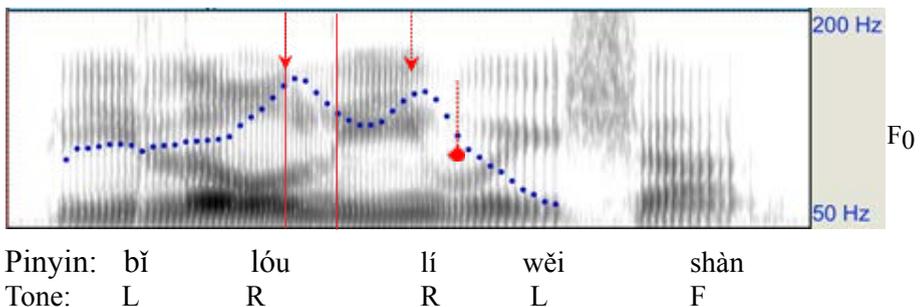
It is not always easy to identify the onset and offset of a movement, however. In the following, we discuss a method that uses a combination of graphical comparison of formant trajectories and  $F_0$ -segment alignment to determine the temporal scope of segments in Mandarin. The first component of the method is minimal contrast comparison of continuous trajectories, which has been applied extensively on  $F_0$  analysis for tone and intonation (e.g., XU, 1999; XU; XU, 2005). For segmental analysis, minimal contrast comparison has been applied in analysis of articulatory data (BOYCE; KRAKOW; BELL-BERTI, 1991; GELFER; BELL-BERTI; HARRIS, 1989), but it has not been widely used in formant analysis, partly because of a lack of convenient tools, which is no longer the case with the availability of FormantPro. The key to minimal contrast comparison of trajectories is to graphically compare the contrasting movements in question in identical or near-identical contexts. This way, aspects of the trajectories that are due to contextual variations are made

identical, so that the differences between the contrasting trajectories become unambiguous.

The second component of the method is to use  $F_0$  events, such as turning points, as temporal anchor points to align the contrasting trajectories. The rationale comes from findings of consistent  $F_0$ -segment alignment in various languages (ARVANITI *et al.*, 1998; LADD *et al.*, 1999; SCHEPMAN *et al.*, 2006; XU, 1998). That is, other things being equal, certain  $F_0$  turning points regularly occur near the onset or offset of a syllable. In Mandarin, for example, the  $F_0$  of the Rising tone (T2) consistently peaks right after syllable offset when followed by a Low or Rising tone. In Figure 5, for example, where the second and third syllable both have the Rising tone, the first  $F_0$  peak occurs right after the onset of the [l] murmur in “lí”.

The significance of the constant  $F_0$ -segment alignment is that it goes both ways. That is, it is also the case that the segmental events involved are likewise aligned to the  $F_0$  events. This further means that  $F_0$  events can be used to determine segmental alignment when there is a lack of landmarks, e.g., in the case of glides and approximants. For example, as found in Xu and Liu (2007), when the  $F_0$  peak is used as the temporal reference, the equivalent of the [l] closure onset in “wěi” would be at the second arrow in Figure 5, as opposed to the low turning point of F2 at the diamond head arrow which has been suggested as a landmark (STEVENS, 2002).

FIGURE 5 – Spectrogram of “Bǐ Lóulí wěishàn” [more hypocritical than Louli], with pitch track (blue speckles) generated by Praat. The two vertical lines mark the onset and offset of [l] closure.



### 3.1 An illustrative experiment

A preliminary experiment was designed to assess the temporal scope of consonants and vowels in CV syllables in Mandarin. One set of the stimuli is shown in Table 1. The stimuli are  $C_1V_1\#C_2V_2$  disyllabic words that form four triplets, each shown in a row in the table. In each triplet, the first two words differ from each other in  $C_2$ : [j] vs. [l], while the second two differ in  $V_2$ : [i] vs [u]. The first two words therefore form a minimal pair for which the divergent point of their F2 trajectories would indicate the onset of  $C_2$ , and the second two words form a minimal pair for which the divergent point of F2 would indicate the onset of  $V_2$ . The two consonants are both sonorants that do not involve full closure of the oral cavity, thus allowing continuous formant movements to be seen during the consonantal constrictions. In addition, all the words have the Rising tone (Tone 2, with the tone mark [ˊ]) on both syllables, so as to allow the occurrence of two  $F_0$  peaks that can serve as time references for the onset and offset of the second syllable.

TABLE 1 – Disyllabic words in 3-way contrasts: [j]/[l] as initial C between words in the first two columns, and [i]/[u] as nuclear V between words in the last two columns.

| Pinyin | Chinese | Pinyin | Chinese | Pinyin | Chinese |
|--------|---------|--------|---------|--------|---------|
| léiyí  | 雷姨      | léilí  | 雷黎      | léilú  | 雷庐      |
| máyí   | 麻姨      | máilí  | 麻黎      | máilú  | 麻庐      |
| lóuyí  | 娄姨      | lóulí  | 娄黎      | lóulú  | 娄庐      |
| lúyí   | 卢姨      | lúlí   | 卢黎      | lúlú   | 卢庐      |

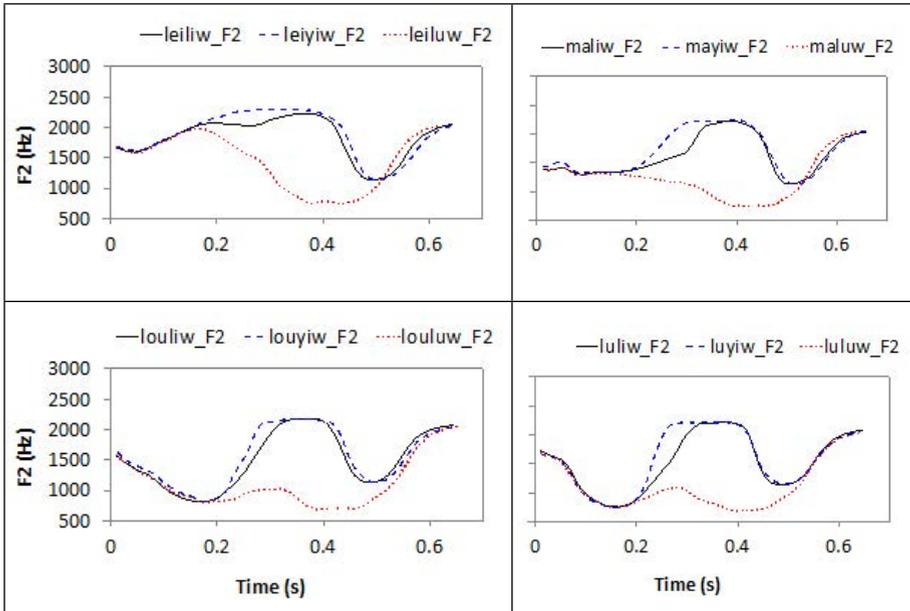
Three male native speakers of Mandarin read aloud the triplets, each in the carrier “Bǐ \_\_\_ wěishàn” [more hypocritical than \_\_\_], with 8 repetitions each, in separate randomized blocks. Their formant trajectories were extracted with FormantPro, and their  $F_0$  patterns with ProsodyPro (XU, 2013). A separate Praat script was written to align the formant trajectories with respect to the  $F_0$  peaks associated with the two Rising tones in each word. All the formant trajectories were taken at 20 evenly spaced locations in each syllable after the  $F_0$ -based boundary adjustment. Mean trajectories were then obtained by averaging across the repetitions as well as speakers. At the same time, time values at each of the 20 points

were also averaged across the repetitions and speakers, which will serve as time axes for some of the formant plots in the analysis.

### **3.2 Graphical analysis and discussion**

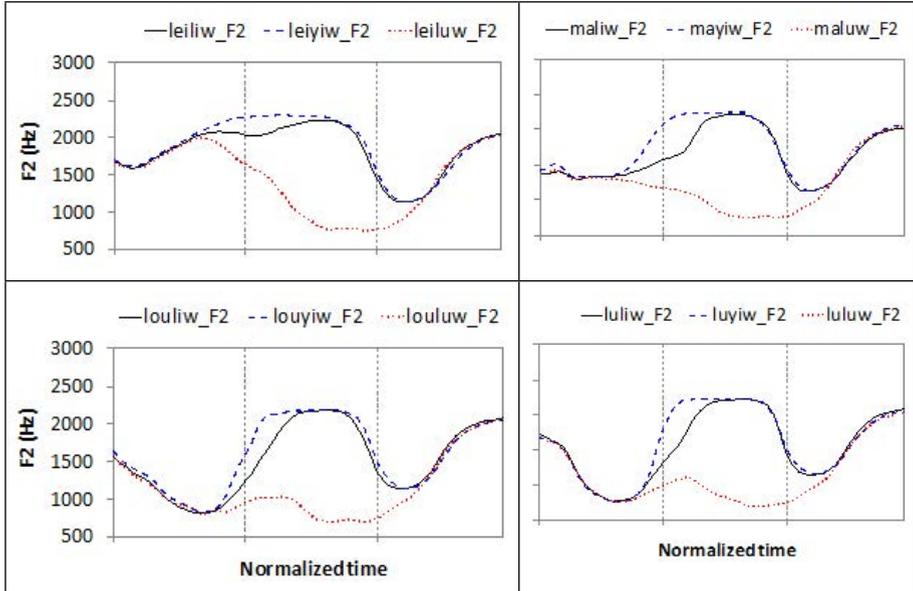
Figure 6 displays grand mean F2 trajectories of the four triplets in Table 1. In each plot, the solid and dashed lines differ in the initial consonants: [l] vs. [j], and the point at which the two trajectories start to diverge would indicate the onset of both consonants, as it is where the articulatory movements start to move toward their respective targets. The solid and dotted lines, on the other hand, differ in the vowels of the second syllable: [i] vs. [u], and the point at which the two trajectories start to diverge would indicate the onset of both vowels. Strikingly, in each case the vowel divergent point occurs at about the same time as the consonant divergent point. Since the contrasting syllables are [li] and [lu], the V approaching movements actually also includes movements toward[l], as revealed by the contrast between [li] and [ji]. In other words, contrary to the conventional view that the acoustic onset of the vowel starts much later—i.e., at the voice onset—than that of the consonant in a CV syllable, the F2 dynamics suggests that the two may actually start at the same time.

FIGURE 6 – Mean F2 trajectories of four triplets in Table 1, plotted on mean time relative to the onset of [l] or [m] in the first syllable of the target word. Both F2 and time are averaged across 8 repetitions by 3 male speakers.



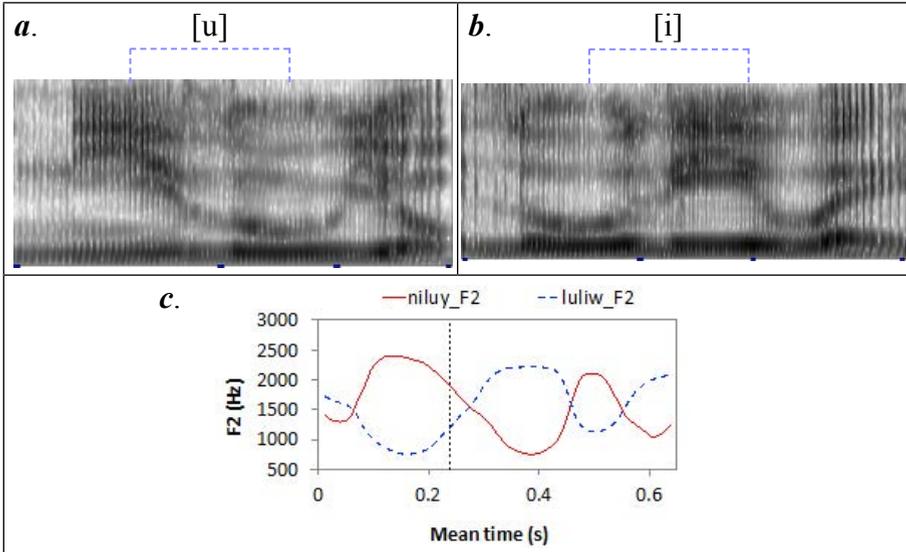
To further explore the exact location of the common starting point of C and V, the  $F_0$ -aligned F2 trajectories are plotted on normalized time in Figure 7. The two vertical lines in each plot are at the  $F_0$  peaks, which divide the formant trajectories into three intervals, each corresponding to one of the conventional syllables. The time-normalized F2 trajectories show greater consistencies within all the triplets than those on averaged real time in Figure 6, indicating the relevance of syllable as a unit of articulatory target approximation. Most relevantly, the joint onset of C2 and V2 movement toward their respective targets can now be seen as well before (about 50–100 ms based on a preliminary estimate) the conventional syllable onset.

FIGURE 7 – Mean time-normalized F2 trajectories of the four triplets in Table 1, averaged across 8 repetitions by 3 male speakers. The two vertical lines in each plot are at the F<sub>0</sub> peaks.



The significance of the new estimate of vowel onset in CV syllables is even more striking when formant movements are spectrally visible across conventional boundaries thanks to the articulatory transparency of [l], as can be seen in Figure 8. In (a), F2 moves continuously from its highest position in the middle of the vocalic section of [ni] to the middle of the vocalic section of [lu]. Assuming that movement toward a target is the scope of a vowel as hypothesized, this entire downward movement would constitute the temporal scope of [u]. Likewise, the entire rising movement of F2 in Figure 8b would constitute the temporal scope of [i]. These scopes are strikingly different from the conventional segmentation as marked by the transcriptions below the spectrograms. When the entire formant trajectories of [li] and lu] are laid on top of each other in Figure 8c, they form a mirror image that makes the temporal domains of the vowels even less ambiguous.

FIGURE 8 – (a, b) Spectrograms of [ni lu jou] and [lu li wei] in Mandarin, with conventional segmentation at the bottom and formant-dynamics-based segmentation of [u] and [i]. (c) Mean F2 trajectories of the two words averaged across 8 repetitions by 3 male speakers, plotted on mean time of all tokens of the two words.



What we have demonstrated above is not entirely new, because Öhman (1966) already reported that articulatory movements toward the nuclear vowel in a CV syllable may start during the intervocalic consonant. Research based on articulatory phonology (BROWMAN; GOLDSTEIN, 1992) and the task dynamic model (SALZMAN; MUNHALL, 1989) has also shown heavy overlap of C and V at the syllable initial position. However, in the widespread common practice, vowels are still routinely assumed to start at the consonant release, and any acoustic properties that may reflect the vowel during or before the initial consonant are attributed to anticipatory coarticulation (FOWLER; SALTZMAN, 1993; LINDBLOM; SUSSMAN, 2012). There may be two reasons for the endurance of the conventional segmentation. Firstly, the landmarks are just too visually compelling to ignore: How can the [u] in Figure 8a, for example, start from where the formants clearly indicate the vowel [i], and by so doing cross the entirety of the consonant [l] right in the middle? Secondly, the articulatory-based segmentation is often auditorily implausible: The [u] and [i] segments as suggested in Figure

7, for example, would both sound like two syllables due to the [l] closure in the middle. How can they be considered as corresponding to single vowels? For an articulatory-based acoustic segmentation to be sufficiently compelling, it is necessary to demonstrate that articulatory dynamics is in fact directly reflected in the acoustics. The direct visualization of continuous formant trajectories like those shown in Figures 6-8 generated by FormantPro allows us to see that articulatory dynamics is actually much more acoustically transparent than is generally believed. Thus there may be sufficient ground to assume articulatory and acoustic segmentations as fundamentally the same.

### 3.3 Caveats

The preliminary data in this section are presented mainly for illustrating the use of FormantPro. The methodology described is designed only for the specific question addressed in the study. In particular, two clarifications are in order. First, the use of  $F_0$  as a reference is only a useful strategy rather than a mandatory requirement for formant analysis. What has been demonstrated is that, like for  $F_0$ , the dynamic aspect of segmental articulation can be studied by examining continuous formant trajectories with the availability of FormantPro, and it is possible to also combine it with ProsodyPro to explore some questions in ways that go beyond what can be done with conventional methods.

Secondly, despite the preliminary evidence for simultaneous onset of consonant and vowels that is much earlier than those based on standard practice, it is not yet clear how the finding, if further confirmed, can be used in phonetic segmentation of speech utterances for annotation purposes. One possibility is to establish, through large-scale empirical testing, segmentation rules that can be easily applied in practice. For example, for simple CV syllables where the C is an obstruent consonant, the C-V co-onset point can be set at a fixed amount of time, e.g., 50 ms (which could be speaker-specific due to individual differences in articulation rate) ahead of the easily observable closure onset.

## 4. Conclusions

In this paper we have introduced FormantPro, a Praat-based research tool for systematic analysis of formants. The tool facilitates analysis of articulatory dynamics through direct comparison of

continuous formant trajectories. This is achieved by, among other things, allowing users to obtain time-normalized formant trajectories that can be averaged across repetitions as well as speakers. It also facilitates systematic analysis of large amount of experimental data by automating procedures that do not require human judgment and saving a variety of formant, duration and intensity measurements in formats that are ready or near-ready for statistical analysis. As an illustration, we also presented preliminary data from a study aimed at assessing the temporal scope of consonants and vowels in CV syllables in Mandarin. These data provide evidence that the temporal scope of vowel is much larger than what is mostly assumed in common practice, as it starts roughly at the same time as the initial consonant. The new evidence for the co-onset of C and V may lead to a new discussion of coarticulation that treats the identification of temporal scope of phonetic units as a prerequisite.

#### **Authors' contribution**

Y.X. developed FormantPro and conceived the experiment. H.G. carried out the experiment. Y.X. and H.G. jointly performed data analysis and co-wrote the paper.

#### **References**

- ARVANITI, A.; LADD, D. R.; MENNEN, I. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, Elsevier, v. 36, p. 3-25, 1998.
- BERKSON, K.; DAVIS, S.; STRICKLER, A. What does incipient/ay/-raising look like?: A response to Josef Fruehwald. *Language*, Washington, v. 93, n. 3, p. e181-e191, 2017.
- BOERSMA, P. Praat, a system for doing phonetics by computer. *Glott International*, Blackwell Publishing, v. 5, n. 9/10, p. 341-345, 2001.
- BOYCE, S. E.; KRAKOW, R. A.; BELL-BERTI, F. Phonological under specification and speech motor organization. *Phonology*, Elsevier, v. 8, p. 210-236, 1991.
- BROWMAN, C. P.; GOLDSTEIN, L. Articulatory phonology: An overview. *Phonetica*, International Society of Phonetic Sciences, v. 49, p. 155-180, 1992. Doi: [10.1159/000261913](https://doi.org/10.1159/000261913)

CHENG, C.; XU, Y. Articulatory limit and extreme segmental reduction in Taiwan Mandarin. *Journal of the Acoustical Society of America*, v. 134, n. 6, p. 4481—4495, 2013.

FOWLER, C. A.; SALTZMAN, E. Coordination and coarticulation in speech production. *Language and Speech*, Sage Journals, v. 36, n. 2-3, p. 171-195, 1993.

GAO, H.; XU, Y. Coarticulation as an epiphenomenon of syllable-synchronized target approximation—Evidence from F0-aligned formant trajectories in Mandarin. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 135, Pt. 2, 2013.

GELFER, C. E.; BELL-BERTI, F.; HARRIS, K. S. Determining the extent of coarticulation: effects of experimental design. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 86, n. 6, p. 2443-2445, 1989.

LADD, D. R.; FAULKNER, D.; FAULKNER, H.; SCHEPMAN, A. Constant “segmental anchoring” of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 106, p. 1543-1554, 1999.

LEE, A.; MOK, P. Acquisition of Japanese quantity contrasts by L1 Cantonese speakers. *Second Language Research*, Hong Kong, 2017. Doi: <http://dx.doi.org/10.1177/0267658317739056>

LINDBLOM, B.; SUSSMAN, H. M. Dissecting coarticulation: How locus equations happen. *Journal of Phonetics*, Elsevier, v. 40, n. 1, p. 1-19, 2012.

LIU, H.; LIANG, J. Vowels as acoustic cues for sub-dialect identification in Chinese. In: INTERNATIONAL SYMPOSIUM CHINESE SPOKEN LANGUAGE PROCESSING (ISCSLP), 10<sup>th</sup>., Tianjin, China, 2016. *Proceedings...* Tianjin, China: IEEE, 2016. p. 1-5.

ÖHMAN, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 39, p. 151-168, 1966.

SALTZMAN, E. L.; MUNHALL, K. G. A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, Francis & Taylor Online, v. 1, p. 333-382, 1989.

SCHEPMAN, A.; LICKLEY, R.; LADD, D. R. Effects of vowel length and “right context” on the alignment of Dutch nuclear accents. *Journal of Phonetics*, Elsevier, v. 34, p. 1-28, 2006.

STEVENS, K. N. *Acoustic Phonetics*. Cambridge, MA: The MIT Press, 1998.

STEVENS, K. N. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 111, p. 1872-1891, 2002.

TURK, A.; NAKAI, S.; SUGAHARA, M. Acoustic Segment Durations in Prosodic Research: A Practical Guide. In: SUDHOFF, S.; LENERTOVA, D.; MEYER, R. *et al. Methods in Empirical Prosody Research*. Berlin; New York: De Gruyter, 2006. p. 1-28.

XU, Y. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, Bankstown, Australia, v. 55, p. 179-203, 1998.

XU, Y. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, Elsevier, v. 27, p. 55-105, 1999.

XU, Y. ProsodyPro — A tool for large-scale systematic prosody analysis. In: TOOLS AND RESOURCES FOR THE ANALYSIS OF SPEECH PROSODY (TRASP 2013), Aix-en-Provence, France, 2013. *Proceedings...* Aix-en-Provence: [s.n.], 2013. p. 7-10.

XU, Y. How often is maximum speed of articulation approached in speech? *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 121, Pt. 2, p. 3199-3140, 2007.

XU, Y.; LIU, F. Determining the temporal interval of segments with the help of F0 contours. *Journal of Phonetics*, Acoustical Society of America, v. 35, p. 398-420, 2007.

XU, Y.; WANG, Q. E. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, Elsevier, v. 33, p. 319-337, 2001.

XU, Y.; XU, C. X. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, Elsevier, v. 33, p. 159-197, 2005.

ZUE, V.; SENEFT, S.; GLASS, J. Speech database development at MIT: TIMIT and beyond. *Speech Communication*, v. 9, n. 3, p. 1-356, 1990.