Beat gestures and prosodic domain marking in French

Patrick L. Rohrer[1,2], Pilar Prieto[3,2] & Elisabeth Delais-Roussarie[2]

[1]Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Catalonia
[2]Laboratoire Linguistique de Nantes (UMR 6310-LLING), Université de Nantes, France
[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Catalonia
patrick.rohrer@upf.edu, pilar.prieto@upf.edu, elisabeth.delais-roussarie@univ-nantes.fr

ABSTRACT

It is widely held that co-speech gestures are produced in a coordinated fashion with prosodic prominence [11, 17]. Studies have shown that gesture strokes and apexes tend to be temporally executed in conjunction with pitch accentuation [8, 15, 21, 6, among others]. Fewer studies have looked at beat gestures as prosodic domain markers, as in French where pitch accentuation can serve a demarcative function on the Accentual Phrase (AP) domain. Prosodic and gestural analyses of an 18-minute long academic-style discourse were carried out, with the goal of exploring the relationship between beat gesture production and prosodic structure in French, where pitch accents have a demarcative function. Our findings show that beat gesture apexes were aligned with pitch accented syllables at much lower rates than previously observed, those that did align prosodically tended to align with AP-final accents, and finally those that did not align tended to occur in AP-initial positions.

Keywords: Beat gesture, prosodic-domain marking, French, Gesture-Speech synchrony

1. INTRODUCTION

1.1. Co-speech Gestures

Co-speech gestures refer to movements of the body (particularly the hands) that are coordinated with speech in terms of timing, semantic, and pragmatic function [16, 17]. Following McNeill's distinction of gesture types [17], we can distinguish referential gestures from non-referential gestures. The former involves gestures where the hands physically resemble the referent in speech (iconic gestures), portray an abstract idea (metaphoric gesture), or refer to the location of an entity in either real or abstract space (pointing or deictic gesture). The latter gesture type, often referred to as beat gestures, are said to be non-referential in that they do not portray lexico-semantic meaning in speech. They are often referred to as beat gestures because of their typical trajectory form of being an up-and-down movement like the conductor of an orchestra. However, this unidimensional view of beat gestures has been challenged [see 22, 20, 21]. Nowadays, beat gestures have been shown to have a more pragmatic function, working to show information structure, discourse-narrative structure, as well as marking rhythm in speech.

It is important to note that McNeill [18] revises his gesture categorization, noting that they should not be considered as discrete boxes into which one gesture can be placed. Instead, he proposes that these "categories" be viewed as dimensions, where gestures may carry properties of various categories.

1.2. Gesture-Prosody Interface

McNeill's original definition of co-speech gestures included the phonological synchrony of gestures with speech. This idea was built on previous research by Kendon [11] who showed a parallel temporal relationship between the organization of the speech stream and of gesture production. Further, Kendon noted that the nuclei of both tone units and gesticular phrases seemed to coincide in time. This led McNeill to establish the phonological synchrony rule, which states that the stroke of a gesture comes before or ends at the phonological peak syllable of speech [17].

Since then, a number of studies have investigated the temporal alignment between gestural and prosodic prominence. One major study by [8] analyzed spontaneous discourse by an American English native speaker at a town-hall meeting, and found that the speaker's gesture apexes co-occurred with a pitch accent 95.7% of the time. In another corpus analysis of American English conversational speech in by [15], the author found that pitch accents and gesture apexes "co-occur repeatedly" (p. 81), and that the distance in milliseconds between the two phenomena were distributed closely around zero, with an average of pitch accents occurring 17 milliseconds after gesture apex. These previous studies, however, did not account for gesture type. A study by [6] investigated the production of deictic gestures by native Catalan speakers in a pointing/naming task and found that gesture apex was tightly correlated with pitch peaks and both were bound by prosodic phrasing.

Specifically regarding non-referential gestures, [21] found that apexes occurred during a pitch-accented syllable 83.13% of the time in a corpus of American English academic-style discourse. More granularly, a study by [13] on the prosodic anchoring of beat gesture during a reading task showed that the closest landmark in speech to beat apexes was the pitch peak in the stressed syllables.

As mentioned, most of these studies used English as the language of study, where pitch accents reflect phrasal-level prominences based on the speaker's intentions. Fewer studies have investigated the co-occurrence of gesture and prosodic prominence in other languages like French, where prominence is fixed at the phrasal level and pitch accents may serve a demarcative function rather than a merely prominence-lending function. In French, pitch accents typically mark the edges of the Accentual Phrase (AP, the smallest prosodic phrase made up of a lexical word and all the functional word that the it governs, see [9, 10, 19, 3]. Indeed, the AP contains an obligatory pitch accent on the right edge (often realized as H*, though not always) and an "optional" rising pitch excursion on the left edge (Hi). The realization of the initial pitch excursion is not entirely clear but it has been suggested for reasons such as to mark the left edge of the AP, to build up rhythmic patterns or to adding emphasis [1, 4, 5]. While the initial accent is not generally prominence-lending, it is sometimes strengthened during emphatic speech [7]. For the purposes of this preliminary study, any initial rise will be referred to as a pitch accent.

To the authors' knowledge, only one study has directly investigated the temporal relationship between beat gestures and prosodic structure in French. Using spontaneous conversational speech, [7] analyzed the co-occurrence of gesture with prosodic emphasis compared to their co-occurrence with thematic structures such as left dislocation and pseudo-cleft constructions. The author describes prosodic emphasis as the presence of an "unusually strong word onset" (p. 2) which would roughly correspond to the (Hi) accent previously described. She showed that gestures reinforced prosodic emphasis more than thematic structures. Particularly regarding prosodic emphasis, she found that beat gestures are associated with prosodic emphasis more than other gesture types. However, the study limits its investigation of prosodic emphasis to initial accents and does not account for the potential effects of phrase-final, obligatory prominence. The current study aims to expand on this previous work, accounting for the temporal relationship between gesture and prosody in French academic-style discourse. The study will determine how gesture production is modulated based on the prosodic structure in French, where pitch accents have a demarcative function of the Accentual Phrase domain. Specifically we ask (a) whether beat gesture apexes in French tend to be temporally associated with pitch accents, as in other languages; (b) whether they occur more often on syllables that are prosodically marked with initial or final accents within the Accentual Phrase, and (c) if they mark prosodic AP phrasing independently of pitch accents.

2. METHODS

2.1. Corpus

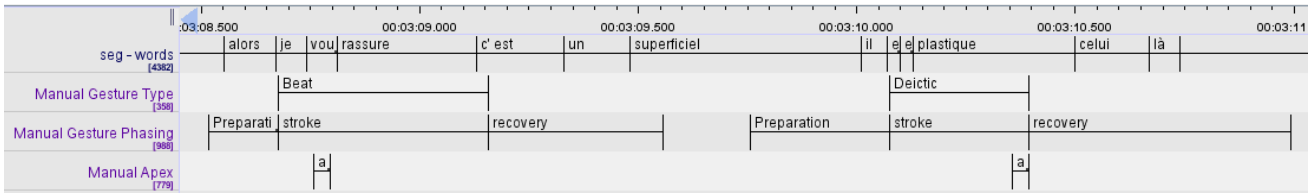
The speech sample used for this study comes from a larger corpus of academic-style or motivational lectures in the form of TED Talks. The TED Talk given by [12] was chosen for (a) his extensive use of gesture, as well as (b) the fact that the video editing of the talk allowed for large stretches of video where his gestures were visible. The original video lasts 18m 24s. Moments where the speaker could not be seen were excluded from the analyses, leaving a total of 11m and 45s for analysis.

2.2. Gestural and prosodic annotation

The gestural annotation was carried out in ELAN [24] by the first author (see Figure 1). The video was annotated for gesture in two phases. First, the temporal structure of gestures was annotated. This was carried out without the audio in order to avoid any potential influences of pitch accentuation on the placement of gesture strokes and apexes. Following the description by [11], gesture units were first annotated and then separated into their individual components (preparation, stroke, recovery, pre-/post-stroke holds). On a separate tier, the apex was annotated.

The apex was determined as follows: for unidirectional strokes, the endpoint was considered the apex, for bidirectional strokes, the point of change in direction was annotated as the apex, and in multidirectional strokes, multiple apexes were annotated at points of zero acceleration [14: 190]. Second, once all gesture phases and apexes were annotated without sounds, the video was replayed with sound to determine the gesture category for each stroke according to [17, 18]. If the gesture showed no clear reference to concurrent speech content in terms of hand form or trajectory, the gesture was labelled as a beat. Ambiguous gesture phasing was also modified or re-categorized in function with speech; however, the temporal boundaries of apexes were not modified.

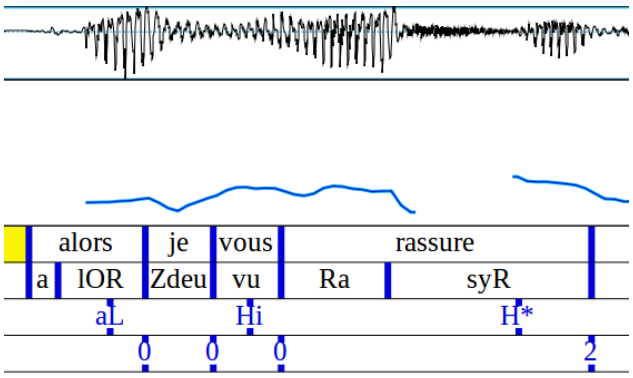
Figure 1: Sample Gesture annotation in ELAN of [12] at 3m 08s showing a beat gesture and a deictic gesture, including tiers for transcription, gesture type, phrasing, and apex.



Prosody was annotated separately using PRAAT [2]. Prosodic annotations followed F_ToBI labelling as laid out in [3] (see Figure 2 below).

Prosodic phrasing is indicated by a "break" tier, where a break level 0 refers to word boundaries that involve clitics. A break level 1 is used to denote boundaries between two lexical words. The smallest prosodic phrase (break level 2) is the Accentual Phrase, which is always bounded by a pitch accent on the right edge which is coded as (T*). An optional rising pitch excursion may occur in the initial syllables of the AP, which is encoded as (Hi). Low tones may occur at the left edge of the AP, labelled as (aL), or before the final pitch accent, labelled as (L). Higher prosodic phrasing was labelled for intermediate (ip) and full intonational (IP) phrases (breaks 3 and 4, respectively), along with their corresponding boundary tones (T- for break 3; T% for break 4).

Figure 2: Sample of F_ToBI annotation corresponding to the beat gesture from Figure 1 at 3m 08s



Once the independent annotations of prosody and gesture were finished, the prosodic annotation was imported into the ELAN file on separate tiers and a CSV file was produced that contained the time stamped data for gesture phrasing, apex position, and the prosodic annotations. This allowed for millisecond precision in determining the alignment between gesture apexes and syllables with initial or final accents.

3. RESULTS

The study will determine how gesture production is modulated based on the prosodic structure in French, specifically asking: (a) whether beat gestures in French tend to be temporally associated with pitch accents, (b) whether they occur more often on initial or final accents within the AP, and (c) if gestures mark prosodic phrasing independently of pitch accents.

Table 1 below shows the total number of gesture apexes produced by gesture type. We found a total of 779 gesture apexes. Of those, 108 were referential in nature (12 iconic, 69 metaphoric, 27 deictic) and 670 were non-referential beat gestures. In other words, 86% of this speaker's gestures were non-referential in nature.

Table 1: The total number of gesture apexes produced per gesture type as per [2].

Gesture Type	Apexes
Grand Total	**779**
Beat	**670**
Referential	**108**
Iconic	12
Metaphoric	69
Deictic	27

Of the 670 beat apexes, a total of 25 apexes were not included in further analyses for two reasons. Nine of them occurred in close temporal proximity to another beat apex, causing them to fall on the same syllable. The other 16 apexes occurred at moments when there was no co-occurring speech. This leaves us with a total of 645 apexes to analyse.

In order to assess the percentage of beat apexes that are aligned with their associated accented syllables, be it initial accents (Hi) or final accents (T*), we counted strict co-occurrence of the apex within the accented syllable. This will be referred to as "strict alignment". However, as apexes that occur up to 200ms before stressed syllables can still be perceived as prosodically aligned [13], apexes that did not strictly occur on accented syllables were further analysed: their distance in milliseconds to the

nearest pitch accented syllable, and the number of syllables from the right edge of the AP were noted. In cases where the non-aligned apex fell within 120 ms of a stressed syllable, the first author revised the video to determine whether perceptually the gesture apex seemed to align with a pitch accented syllable. This distance was chosen as it was just short of our speaker's average syllable duration (148 ms). If the author perceived alignment, the apex was considered as perceptively aligned. These results will be referred to as "non-strict alignment".

Table 2 shows the percentage of strictly and non-strictly aligned beat apexes in regards with respect to initial or final accents. The results show that 56.11% of the beat gestures strictly aligned with a pitch accented syllable. Of these beats that aligned prosodically, 9.76% fell on initial accents and 45.58% on final accents. As for the non-strict alignment, we found that 12.87% were associated with initial accents, 59.22% with final accents, and 27.91% were not loosely aligned with a pitch accented syllable. Chi-squared analyses confirm that beat gestures that coincide with pitch accentuation are more likely to occur on final accents than initial ones (strict alignment, $\chi^2(1, N=357) = 8.41$, $p=0.0037$; non-strict alignment, $\chi^2(1, N=465) = 10.29$, $p = 0.0013$).

Table 2: The percent of alignment of beat gesture apexes with either initial accent (Hi) or final accent (T*) accented syllables in either a strict or non-strict sense. The total number of apexes is given in parentheses.

Accent type	Percent (number) of beats	
	Strict alignment	**Non-strict alignment**
Initial accent (Hi)	**9.76%** (63)	**12.87%** (83)
Final accent (T*)	**45.58%** (294)	**59.22%** (382)
No alignment	**44.65%** (288)	**27.91%** (180)

Finally, it was decided that if the 180 misaligned apexes still occurred within the first 3 syllables of an AP, that it could be considered as marking the left edge of that prosodic phrase [23]. Of the 180 misaligned beat apexes, over 72.23% of the apexes occurred in an AP-initial positions. Specifically, 64 (35.56%) occurred on the first syllable of the AP, 66 (36.67%) on the second syllable, 31 (17.22%) on the third syllable, and only 19 occurring beyond the third syllable. This indeed shows that a good percentage of beat gestures are aligned with initial positions of the AP that contain no pitch accentuation. A two-sample Kolmogorov-Smirnov test showed that this distribution does not significantly differ from that of the initial accent (with or without beats), where 75 (25.33%) occurred on the first syllable, 130 (43.91%) on the second, 60 (20.27 %) on the third, and only 31 (10.49%) occurring beyond the third ($D(5,6) = 0.4$, $p = 0.652$).

4. DISCUSSION

In general terms, the results of this analysis lend support to the hypothesis that academic style discourse is primarily accompanied by non-referential beat gesture [21]. Indeed, the French speaker used non-referential beat gestures 86% of the time. In regards to our first research question, however, it seems that beat gestures are not as tightly associated with pitch accent as in other languages (namely, English). The same study by [21] found rates of association for similar speech styles as high as 83.13%. This is not the case for French. We found much lower rates, especially when considering the data for strict overlap. Even looking at non-strict overlap, we find that only about 73% of the apexes co-occur with a pitch accented syllable. This finding suggests that factors apart from pitch accentuation may be influencing their planning and/or production.

Regarding our second research question, beat gestures seem to co-occur more frequently on AP-final pitch accents. This is a particularly surprising finding as both beat gestures and initial accents can be used to mark emphasis. Perhaps the acoustic features of the final accent (namely having a longer duration) may act as an attractor. It may thus be fruitful to look towards higher levels in the prosodic hierarchy, where nuclear AP-final pitch accents (the last accent in an ip or IP) are accompanied by greater amounts of syllable lengthening [1]. Further, future studies should distinguish initial accent type.

Finally, our findings suggest that beat gestures may be marking the left edge of the AP independently of pitch accents. While the prospect of a prosodic-domain marking function of beat gestures is quite interesting, further exploration needs to be carried out in order to rule out other possibilities. It may be of more interest to look at the syllables where these misaligned beats fall, as they may still be metrically strong syllables, thus marking rhythm in French. Other influences may not be related to speech prosody, such as pragmatic intent. Taken together, these findings suggest that beat gestures are not merely a reflection of rhythmic prominence in the hands, but may well have other prosodic functions such as prosodic-edge marking that warrant further investigation.

5. REFERENCES

[1] Astésano, C., Bard, E.G. & Turk, A. 2007. Structural influences on initial accent placement in French. *Language and speech*. 50(3), 423–46.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International.* 5(9/10), 341-345.

[3] Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S.A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R., Yoo, H.Y. 2015. Intonational phonology of French: Developing a ToBI system for French. In: Frota, S., Prieto, P. (eds), *Intonation in Romance*. Oxford.

[4] Di Cristo, A. 1999. Vers une modélisation de l'accentuation du français (première partie). *Journal of French language studies*. 9(2), 143–179.

[5] Di Cristo, A., 2000. Vers une modélisation de l'accentuation du français (seconde partie). *Journal of French language studies*. 10(1), 27–44.

[6] Esteve-Gibert, N., Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research.* 56(3), 850-864.

[7] Ferré, G. 2014. A multimodal approach to markedness in spoken French. *Speech Communication*. 57, 268-282.

[8] Jannedy, S., Mendoza-Denton, N. 2005. Structuring information through gesture and intonation. *Interdisciplinary studies on information structure*. 3, 199–244.

[9] Jun, S.A., Fougeron, C. 1995. The Accentual Phrase and the Prosodic structure of French. *Proc 13th International Congress of Phonetic Science*, 2, 722-725.

[10] Jun, S.A. & Fougeron, C. 2000. A phonological model of French intonation. In: A. Botinis, (ed), *Intonation: Analysis, Modeling and Technology*. Dordrecht: Kluwer Academic Publishers, 209–242.

[11] Kendon, A. 1980. Gesticulation and speech: Two aspects of the process of utterance. In: Key, M.R. (ed), *The Relationship of Verbal and Nonverbal Communication*. The Hague: Mouton, 202 - 227.

[12] Laroche, D. 2016. "Et si votre rêve devenait possible…" [online] TEDx talk available at https://www.youtube.com/watch?v=fbGfe78_2jc

[13] Leonard, T. & Cummins, F. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes*. 26(10), 1457–1471.

[14] Loehr, D. 2007. Aspects of rhythm and gesture in speech. *Gesture*. 7(2), 179-214.

[15] Loehr, D. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology.* 3(1), 71-89.

[16] McNeill, D. 1985. So you think gestures are nonverbal? *Psychological review*. 92(3), 350–371.

[17] McNeill, D. 1992. *Hand and mind: What gestures reveal about thought.* Chicago: Chicago University Press.

[18] McNeill, D. 2006. Gesture: a psycholinguistic approach. *The encyclopedia of language and linguistics*. 58–66.

[19] Post, B. 2000. Tonal annd Phrasal structures in French Intonation. The Hague: Thesis.

[20] Prieto, P., Cravotta, A., Kushch, O., Rohrer, P.L. Vilà-Giménez, I. 2018. Deconstructing beat gestures: a labelling proposal. In *Proc. 9th International Conference on Speech Prosody 2018*. 201–205.

[21] Shattuck-Hufnagel, S., Ren, A. 2018 (in press). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*.

[22] Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., Demuth, K. 2016. Non-referential gestures in adult and child speech: Are they prosodic? *Speech Prosody 2016*. 836–839.

[23] Welby, P. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics.* 34(3), 343–371.

[24] Wittenburg, P., Brugman, H., Russel, A. Klassmann, A. Sloetjes, H. 2006. Elan: a professional framework for multimodality research. In *Proceedings of LREC*.