PROSODIC STRUCTURE SHAPES THE TEMPORAL REALIZATION OF INTONATION AND MANUAL GESTURE MOVEMENTS

Núria Esteve-Gibert[1], Pilar Prieto[2]

[1] *Universitat Pompeu Fabra, Spain*

[2] *ICREA - Universitat Pompeu Fabra, Spain*

0. ABSTRACT

Purpose: Previous work on the temporal coordination between gesture and speech found that the prominence in gesture coordinates with speech prominence. This study investigates the anchoring regions in speech and pointing gesture aligning with each other. Our hypotheses are that (a) in contrastive focus conditions the gesture apex is anchored in the intonation peak, (b) the upcoming prosodic boundary influences the timing of gesture and intonation movements.

Method: 15 Catalan speakers pointed at a screen while pronouncing a target word with different metrical patterns in a contrastive focus condition and followed by a phrase boundary. A total of 702 co-speech deictic gestures were acoustically and gesturally analyzed.

Results: Intonation peaks and gesture apexes showed parallel behavior with respect to their position within the accented syllable: they occurred at the end of the accented syllable in non-phrase-final position, whereas they occurred well before the end of the accented syllable in phrase-final position. Crucially, the position of intonation peaks and gesture apexes was correlated and was bound by prosodic structure.

Conclusions: The results refine the phonological synchronization rule (McNeill, 1992), showing that gesture apexes are anchored in intonation peaks, and that gesture and prosodic movements are bound by prosodic phrasing.

1. INTRODUCTION

Studies on the relation between gestures and speech in human communication in recent years have shown that gesture and speech are tightly integrated. In his groundbreaking book, McNeill (1992) gives five main reasons to argue that gesture and speech form a single system in communication: (1) gestures occur with speech in 90% of cases, (2) gesture and speech are phonologically synchronous, (3) gesture and speech are semantically and pragmatically co-expressive, (4) gesture and speech develop together in children, and (5) gesture and speech break down together in aphasia. According to McNeill (1992), gesture and speech synchronize at three different levels: at a semantic level, since gesture and speech present the same meanings at the same time, at a pragmatic level, since the two modalities perform the same pragmatic functions when co-occurring, and at a phonological level, as the most prominent part of the gesture seems to be integrated into the phonology of the utterance.

Broadly speaking, studies on the alignment between gestures and speech assume that gesture and speech are temporally aligned in the sense that the most prominent part of the gesture and the most prominent part of the speech co-occur in time (Birdwhistell, 1952; 1970; Kendon, 1972; 1980). Yet, the specific way this temporal alignment operates is still unclear. To date, there are few empirical studies that have investigated this alignment and they have in fact used distinct methodologies: for some studies, for instance, the prominent part of the gesture aligning with speech is the stroke phase, i.e. the interval of time in which the peak of effort in the gesture occurs (McNeill, 1992), whereas for other studies what temporally coordinates with speech is the apex, i.e. the point in time in which the movement reaches its kinetic ‘goal’ (Loehr, 2004). These different methodologies will be examined below.

Similarly, there is no consensus in the literature on what part of the speech aligns with the most meaningful part of the gesture (i.e. the stroke phase or the apex). Previous studies can be separated into four groups according to the conclusion that they draw. The first group concludes that the prominence in gesture coordinates with the focused word (Butterworth & Beattie, 1978; Roustan & Dohen, 2010); the second, that the prominence in gesture coordinates with the lexically stressed syllable (Loehr, 2004, 2007; Rochet-Capellan, Laboissière, Galván & Schwartz, 2008); the third, that the prominence in gesture coordinates with syllables with intonation peaks (Nobe, 1996; De Ruiter, 1998 – *second experiment*); and the fourth, that there is in fact no alignment between the prominence in gesture and speech prominence (De Ruiter, 1998 – *first experiment*; McClave, 1994; Rusiewicz, 2010). They have all used different methodologies, so it is worth taking a closer look at them in order to better understand their results.

Butterworth and Beattie (1978) studied iconic gestures and found that gestures tend to start in a pause just before a corresponding noun, verb, adverb, or adjective. Similarly, Roustan and Dohen (2010) also found that the apex coordinates with prosodic focus in a broad sense. They tested 10 adult French speakers in an experimental task where they had to point at an object appearing on a monitor screen or perform a beat gesture while pronouncing a word in a contrastive focus condition. Their results showed that the apex of pointing and beat gestures overlapped with the focused word.

Another group of results showed that the most meaningful part of the gesture coordinates with prominent (or pitch-accented/stressed) syllables of speech. First in his dissertation and later on in an article, Loehr (2004, 2007) investigated the alignment between gesture strokes and pitch-accented syllables in 15 English speakers while engaged in natural conversations. The author selected four clips from these conversations and studied all the gestures they produced (iconics, deictics, metaphorics, emblems, beats, and head movements). Results showed that gesture apexes were significantly more likely to align with pitch-accented than with non-pitch-accented syllables. Two more studies have investigated the gesture-syllable alignment in experimental settings. Rochet-Capellan et al. (2008) investigated the temporal coordination between deictic pointing gestures and jaw movements in a task where 20 Brazilian Portuguese speakers had to point at a target on a screen while naming a non-word. The alignment of two measures was analyzed: the *pointing plateau* (the amount of time during which the finger remained pointed at the target) and the jaw opening apex. Results showed that when the stress was in the first syllable position (e.g. *'papa*), the beginning of the *pointing plateau* aligned with the jaw opening apex of stressed syllables. However, when the stress was in the second syllable position (e.g. *pa'pa*), the end rather than the beginning of the *pointing plateau* synchronized with the jaw-opening apex.

The coordination between gesture and speech has also been studied focusing on the intonational-metrical structure of speech. In general, these studies have shown that prominent syllables with intonation peaks coordinate with gestural strokes or apexes. Nobe (1996, *experiment 4*) studied iconic, deictic, and metaphoric gestures in cartoon narrations of six English speakers. The authors selected a sample of data (no explanation was given regarding the selection criteria) and, after analyzing it, they found that most of the gestures occurred in speech fragments including both pitch peak and intensity peak. A more precise study was carried out by De Ruiter (1998. *experiment 2*), who tested eight Dutch speakers in a controlled experimental setting in order to investigate the alignment of deictic gestures and prosodic structures in a context of contrastive focus production. Results revealed that prosodic structure influenced the gesture movement, since the gesture launch (i.e. the time between the beginning of the preparation phase and the apex) was longer when the word had the stress in phrase-final position, and shorter when the word had the stress in non-phrase-final position. He thus showed that the apex aligns with the accented syllable irrespectively of the metrical structure of the target word. The author concluded that the contrastive focus condition was crucial to obtain these results because it enhanced the phonetic realization of stress and it introduced a more marked intonational contour in the production of speech.

Such findings were not obtained, however, in the first experiment in De Ruiter (1998). The author analyzed the coordination between deictic pointing gestures and lexically stressed syllables (and not pitch-accented syllables as in his second experiment) in disyllabic and trisyllabic words with different metrical patterns. Participants were eight Dutch speakers who were presented with a controlled setting in which they had to point at a target picture on a screen while naming it. His results showed that the temporal parameters of the gesture apex were not affected by stress position. Against his previous hypotheses, he actually found that gesture apexes did not always occur in stressed syllables. Thus, when the stress was in phrase-final position, apexes occurred in non-phrase-final position, and vice-versa. Therefore, not all studies investigating the coordination between the prominent part of gestures and prominent syllables have found that they are synchronized. Actually, the first author to study concrete examples of gesture-speech combinations (McClave, 1994) did not find temporal coordination between the meaningful part of the gesture and stressed syllables either. This author studied beat gestures in spontaneous conversations of four speakers and her quantitative results revealed that apexes of beat gestures coincided with both stressed and unstressed syllables. More recently, Rusiewicz (2010) analyzed the influence of contrastive pitch accents and syllable position on the apexes of deictic pointing gestures. 15 English speakers participated in a controlled task where they had to point at a target object on a screen while naming it. The results of this study showed that metrical structure did not influence the timing of the gesture, since gesture apexes were time-aligned with the initial onset of the target word rather than occurring with prosodically prominent syllables.

Thus, results of most of the previous studies have shown that in multimodal communication the most prominent part of gestures coordinate with prominent syllables, whether lexical stressed syllables or pitch-accented syllables. However, the coordination of gesture and speech has been studied rather imprecisely, by comparing two distinct measures: some have investigated alignment of two intervals of time, namely the stroke of the gesture and the prominent syllable or prominent word; and others have investigated the alignment between one specific moment and an interval of time, namely the apex of the gesture and the most prominent syllable or word. However, as far as we know, no study has tried to investigate the alignment between two specific points in time: the apex of the gesture and a specific point within the most prominent syllable. Only Rusiewicz (2010) took the vowel midpoint as a measure to compare its alignment with the apex of the gesture. The vowel midpoint is nevertheless an arbitrary anchoring point in the syllable with no relation to its intonation-metrical structure. Instead of the vowel midpoint measure, we propose another anchoring point that may attract the prominence in gesture, namely the peak of the fundamental frequency line, the intonation peak. To our knowledge, by investigating the position of the pitch peak in relation to the peak of the gesture (i.e. the apex), our study will be the first to take two specific points in time as measures for studying the synchrony between gesture and speech.

The goal of this study is to investigate how gesture and speech structures synchronize in time. In order to investigate this issue, the timing of co-speech deictic gestures will be analyzed with respect to the prosodic structure of target words and sentences. Crucially, the target words will have distinct metrical structures and will be produced in a contrastive focus condition. In Catalan, contrastive focus is produced with fronting of the focused word, which is produced with a prominent L+H* pitch accent and crucially followed by an L- boundary tone, which will allow us to trigger distinct stress and intonation peak positions (see section 2.3 for further details). We will test two hypotheses, namely: (a) that it is the peak of the focused pitch accent in the intonational structure that serves as an anchoring site for the most prominent part of gestures; and (b) that both intonation and gestural movements will be bound by prosodic structure, and thus the presence of the upcoming boundary L tone will produce similar effects in both intonation and gestural movements. From the analyses, two potential conclusions can be obtained: that the gesture apexes coordinate with the accented syllable (if they occur during the accented syllable but their position with respect of the accented syllable does not vary when the accented syllable change its position in the intonational phrase), or that the gesture apexes coordinate with F0 peaks and show the same alignment patterns (if their position varies depending on the position of the accented syllable with respect to the phrase boundary).

The present study will focus on the temporal coordination of co-speech pointing deictic gestures. Pointing deictic gestures have been chosen because contradictory results have been obtained when studying the synchronization between gesture and speech in this gesture type (De Ruiter, 1998; Loehr, 2004, 2007; Rusiewicz, 2010). Our aim is to contribute with more evidence to the previous work on the temporal coordination of gesture and speech prominence.

2. METHOD

2.1. Participants

The participants were 15 adult native speakers of Catalan (2 men, 13 women) aged 20-30 years. They were paid €5 for participating in the study and were all right-handed. They were not aware of the purpose of the experiment.

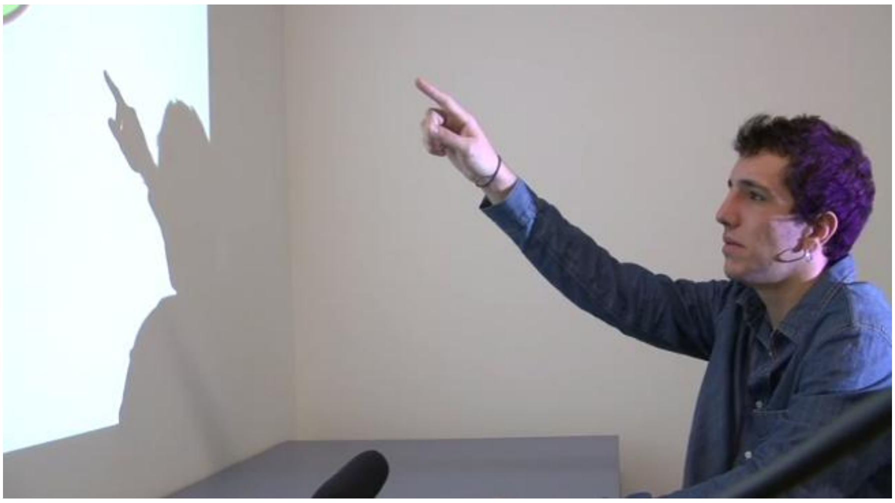
2.2. Procedure

Participants were audio-visually recorded using a Panasonic HD AVCCAM, which can record at 25 frames per second. The camera was placed at a distance of approximately one meter from the participant, as close as possible to capture the entire pointing arm movement, and focused on the participant's left profile. Also, a RØDE NTG-2 microphone was placed 20 cm distance from the participant's mouth.

The procedure we followed was based on Rochet-Capellan et al. (2008).[1] Each participant was seated at a table that was approximately 50 cm from the screen (Figure 1 shows the experimental setting). Participants had to perform a pointing-naming task in which they had to point at a smiley face projected on the screen. The projected face was first red and after a short period of time (varying from 1 second to 3.5 seconds), the red face turned into green. At that specific point in time participants had to pronounce a sentence projected right under the colored face. This color changing worked as a 'go signal', and was performed to avoid anticipatory responses and to make participants pay attention to the experiment. A black square pasted on the middle of the table was used to indicate the finger resting position. An experimental setting and not a naturalistic conversation was chosen to test our hypothesis because the pointing-naming task permitted us to control with high accuracy two sets of variables that were crucial for our purpose, i.e. arm trajectory with respect to prosodic structure. Also, other studies in the field (De Ruiter, 1998; Rusiewicz, 2010) used a methodology consisting of an experimental pointing-naming task, so we would be able to compare our results with theirs.

Figure 1. Experimental setting.

[1] *Our procedure was based on Rochet-Capellan et al. (2008) because we elicited the pointing-naming gestures following their experimental design and procedure. However, our data recording and post-processing methodology differed from theirs, since they used Optotrak for their analysis of arm and jaw movements while we used acoustic and visual measures. See section 2.4.*



Prior to the experiments, participants were told to imagine that the instructor had just pointed at an object on the screen and pronounced the name of that object incorrectly. Then participants were told that they now had to point at the smiley face to indicate its exact position to the instructor while pronouncing its name properly. This pointing-naming task allowed us to obtain a co-speech deictic gesture with a target word focalized both with respect to the speech and the gesture: the target word was pronounced in a contrastive focus condition, and the pointing gesture focally served to establish joint attention and to inform about the location (Enfield, Kita & De Ruiter, 2007). All the information relative to how the instructor mispronounced the object and its correct pronunciation appeared below the smiley face in a sentence such as ‘papa, es diu, i no papà’ (*It’s called Papa, not Papà)*’. Thus, participants had to point at the smiley face while pronouncing the focus word *Papa*. The accented syllables of the target words were underlined to facilitate the reading task. These instructions also appeared on alternate slides during the experimental trials. In addition, before starting with the experiment, participants were briefly trained to become familiarized with the task: first, they had to pronounce aloud the words that would appear during the experiment, and second, they had to practice the pointing-naming task they would perform.

The experiment was divided into three blocks. The first block was a familiarization task in which participants were asked to read aloud the sentences containing the target words they would encounter during the experimental trial. The second block was also a familiarization task in which participants had to practice the pointing-naming task with one repetition for each particular stimulus (12 items in total). Finally, the third block consisted of the experimental trial, where 4 repetitions per stimuli were randomly displayed on the screen (12 items x 4 repetitions = 48 tokens).

During the second block, in which participants had to practice the pointing-naming task, it could happen that participants pointed at the screen while naming other parts of the sentence but not the target word. In such cases, they were reminded of the purpose of their pointing with the following instruction: “Remember that you are correcting my pronunciation of the target word while pointing to its exact position. The important information is the proper name ‘mama’ (for instance), so the rest of the sentence is as if you are telling this information to yourself”. Once this reinforcement of the instructions was given, the second block with the familiarization of the pointing-naming task was run again. Nevertheless, only two participants out of the 15 that participated in the study had to repeat this second block with the familiarization materials.

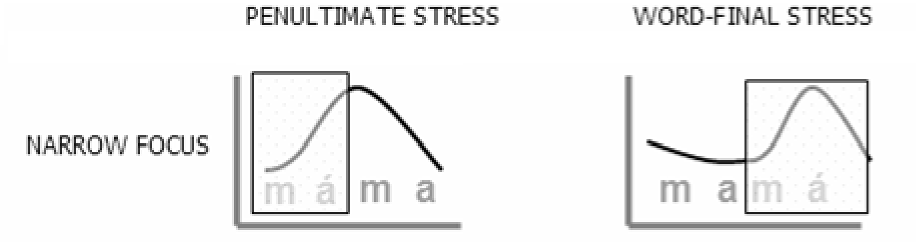
2.3. Materials

The target words in focus appearing below the smiley face had varying numbers of syllables and varying stress positions, namely 'CV, CV'CV, and 'CVCV. The target consonants also varied to keep participants’ attention focused on the task. Thus, the following twelve target words were used: three monosyllables (['pa], ['ta], ['ma], ['na]), three trochees (['papə], ['tatə], ['mamə], ['nanə]), and three iambs ([pə'pa], [tə'ta], [mə'ma], [nə'na]). These test materials included voiceless consonants surrounding some of the target accented vowels (as in the words ['papə], ['tatə], etc.). This was not a problem for the location of the F0 peaks, since as mentioned by previous studies on the location of F0 peaks in contrastive focus pitch accents (L+H*) in Catalan (Vanrell, 2011), F0 peaks in contrastive focus pitch accents are found to be reliably located at the end of the accented vowel.

The target words with different metrical patterns were pronounced in a contrastive focus condition and were followed by a clear phrase boundary (induced by the semantic meaning of the utterance and by an orthographic comma) to trigger different temporal positions of the intonation peaks within the accented syllables (see figure 4).

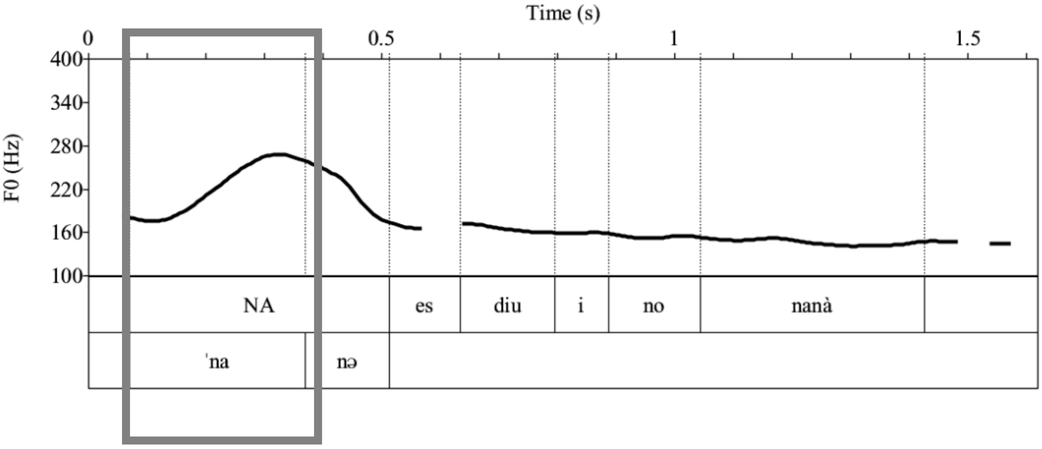
It is well established that F0 peak location varies with the presence of upcoming prosodic events, such as other pitch accents (clash situations) and boundary tones (for a summary, see Silverman & Pierrehumbert, 1990). There is a large body of cross-linguistic research on tonal compression effects due to tonal crowding which have demonstrated that pitch pressure environments (namely, proximity to a boundary tone or to an upcoming pitch accent) can drastically affect surface H alignment patterns (Prieto, 2005 for Catalan; Prieto, van Santen & Hirschberg, 1995 for Spanish; Silverman & Pierrehumbert, 1990). Specifically, the location of F0 peaks in rising pitch accents (namely L+H*) can be directly influenced by the presence of an upcoming low boundary tone. When H* accents occur in phrase-final nuclear position before L% boundary tones they are moved to the left by the closeness of the boundary tone (see work on rising accents on English by Silverman & Pierrehumbert, 1990, on Spanish by Prieto et al., 1995, and on Italian by D'Imperio, 2001). That is, the articulatory gesture for the pitch accent is repelled earlier in time in order for the pitch accent and the following boundary tone to be produced in the same tone bearing unit (the phrase-final syllable). Figure 2 displays a schematic representation of the distinct locations of the F0 peak in the two metrical patterns and in a narrow/contrastive focus condition and Figure 3 shows three examples as a comparison of the F0 movement produced by participants in the contrastive focus conditions when the stress was in the word-final syllable (central and bottom panel) and when the stress was in the penultimate syllable (top panel).

Figure 2. Schematic representation of the realization of the nuclear pitch accent in narrow contrastive focus utterances, in words with penultimate stress and word-final stress (Prieto & Ortega-Llebaria, 2009).



The contrastive focus utterances produced during the pointing-naming task followed the expected prosodic pattern (see Vanrell 2011) (see Figure 4). Target sentences consisted of two prosodic phrases. The first prosodic phrase contained the fronted focus of the communicative act, i.e., the highlighted element with sentence-level prominence, which was pronounced with a prominent L+H* pitch accent on the accented syllable, followed by an L- boundary tone. And the second prosodic phrase was the post-focal material with prosodic deaccenting, i.e. produced with global pitch range compression. Two examples of these sentences can be observed in Figure 4. In them, we can appreciate the location of the nuclear L+H* pitch accent over the accented syllable of the focal elements (NAna in the top panel and naNÀ and NÀ in the central and bottom panels respectively). The difference between the three realizations of the focal prominence is the fact that the L- boundary tone has to be realized within the accented syllable in the iambic and monosyllabic conditions (e.g., naNÀ and NÀ, central and bottom panels), whereas the L- boundary tone can be realized in the posttonic syllable in the trochaic condition (e.g., Nana, top panel)..

Figure 3. Fundamental frequency of the target sentence pronounced by one participant in the three metrical structures: trochaic condition (top panel), iambic condition (central panel), and monosyllabic condition (bottom panel).



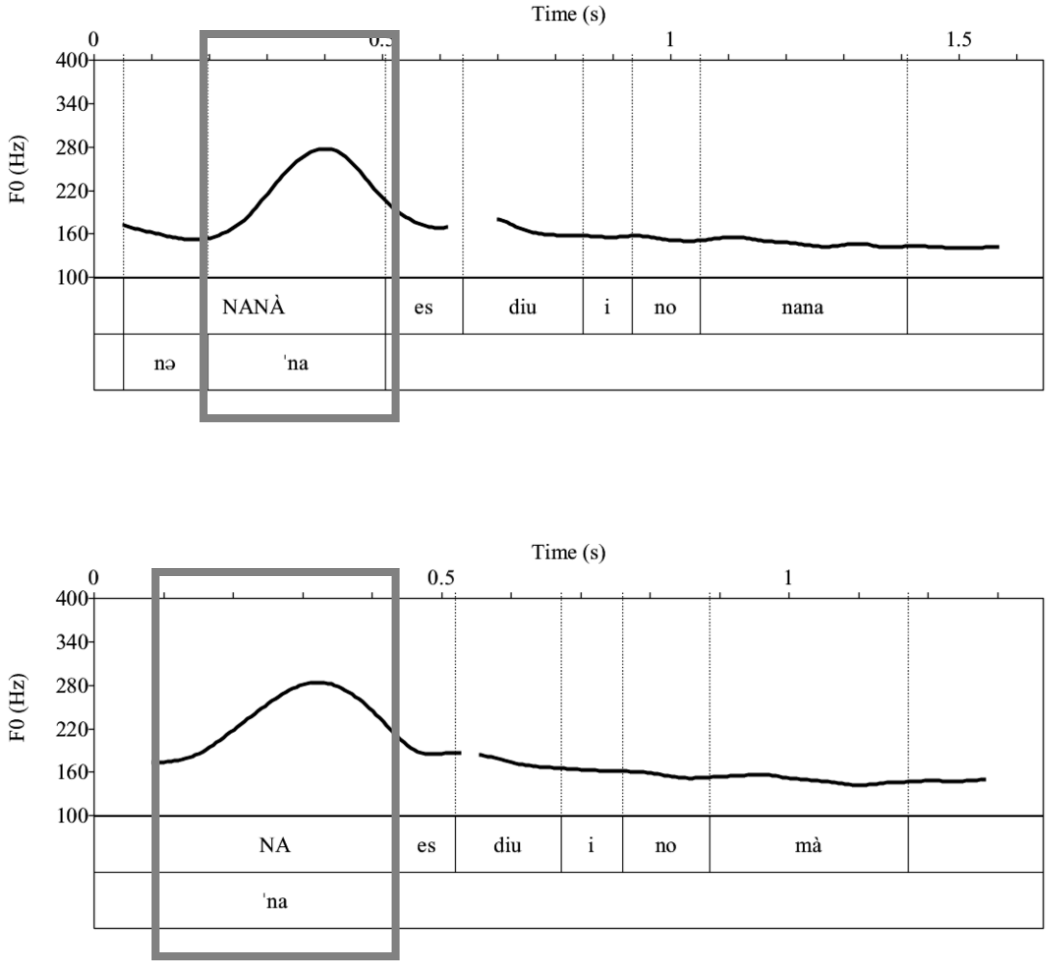
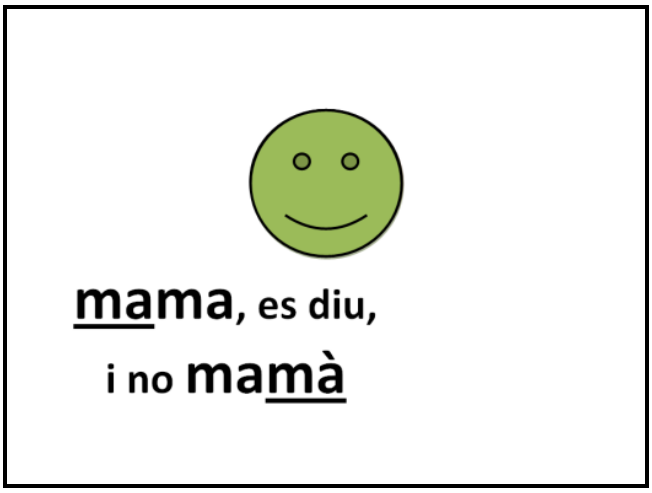


Figure 4. Example of a slide projected in front of the participants with the sentence *mama, es diu, i no mamà* (in English, 'mama, is called, and not mamà').



<u>ma</u>ma, es diu,

i no **ma<u>mà</u>**

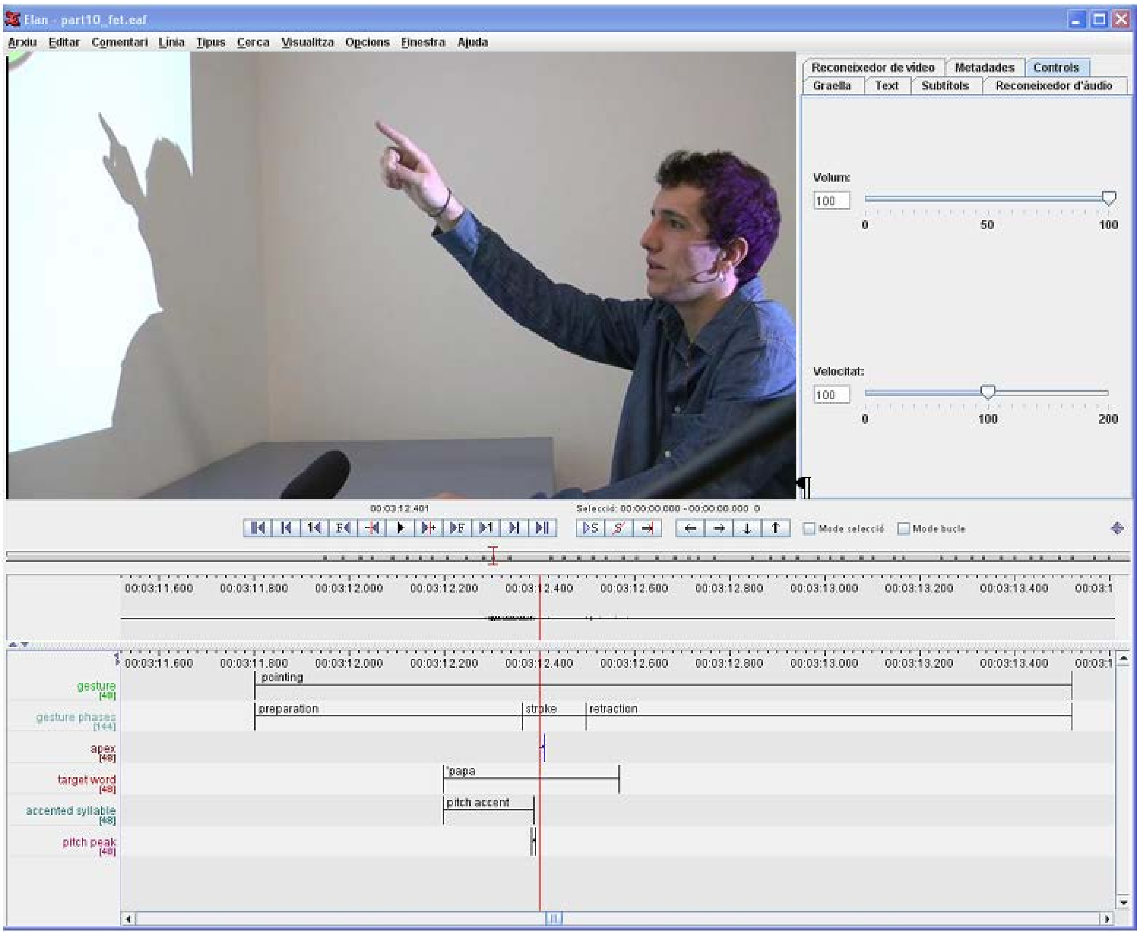
2.4. Coding

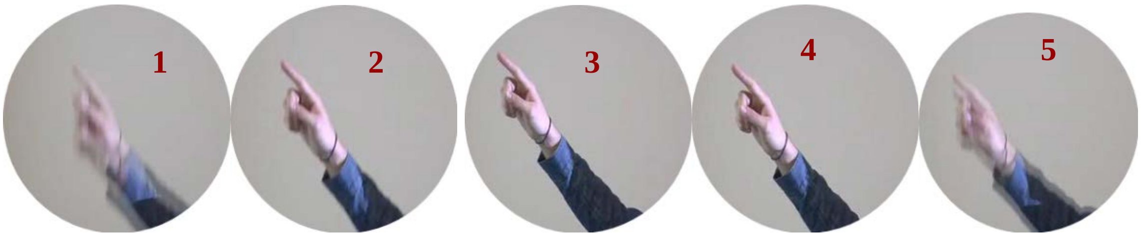
In total, 720 instances of deictic gesture-speech combinations were obtained from the experimental task. From these, 17 gesture-speech combinations were excluded from the analysis because the target words were mispronounced (16 items) or because the pointing movement was interrupted (1 item). Thus, 703 instances of deictic gesture-speech combinations were acoustically and gesturally analyzed.

The gesture and acoustic measures were annotated using ELAN software package (Lausberg & Sloetjes, 2009) in separate tiers, as shown in Figure 5. As for the gesture analysis, three distinct measures were coded. First, we measured the limits of the pointing gesture, from the start of the preparation phase (i.e. the initial arm movement) to the end of the retraction phase (i.e. the final arm movement) (tier 1 in Figure 5). The second measurement involved gesture phases (tier 2 in Figure 5): following McNeill (1992), the gesture phases taken into account were the preparation (the arm move from the rest position until the peak of effort in the gesture), the stroke (the peak of effort in the gesture), and the retraction (the arm moves from its farthest extension to the rest position again). Third, we measured the apex (tier 3 in Figure 5), which is a specific point in time within the most meaningful part of the gesture. The apex measurement has been used in other studies investigating the gesture-speech synchronization because it allows us to identify the point of maximum prominence in the gesture.

In order to locate the stroke and apex of the pointing gesture, we examined the video file. ELAN allows precise navigation through the video recording, i.e., frame by frame. Even though he software program also allows an even more precise annotation (2 ms by 2 ms), this option could not be applied because the video was recorded with a frame rate of 25 frames per second. From the one hand, the stroke of the gesture was annotated in those video frames in which the arm was well extended with no blurring of the image, the fingertip being or not totally extended. Despite the absence of blurring of the image, the arm was not totally static during the interval of the gesture stroke, with the fingertip moving a few pixels back and forth. From the other hand, the gesture apex was annotated in the specific video frame in which we located the furthest spatial excursion of the fingertip during the interval of time in which the arm was maximally extended. When participants performed the pointing gesture slower, this point in time of gesture peak could last more than one frame (normally two frames). In those cases, the gesture peak was considered to be the last of these video frames.

Figure 5. Image stills of a pointing gesture annotated in ELAN. Below, five specific enlarged images during the gesture: (1) during the preparation phase, (2) during the stroke phase and before the apex is reached, (3) at the apex, (4) during the stroke and after the apex is reached, and (5) during the retraction phase.

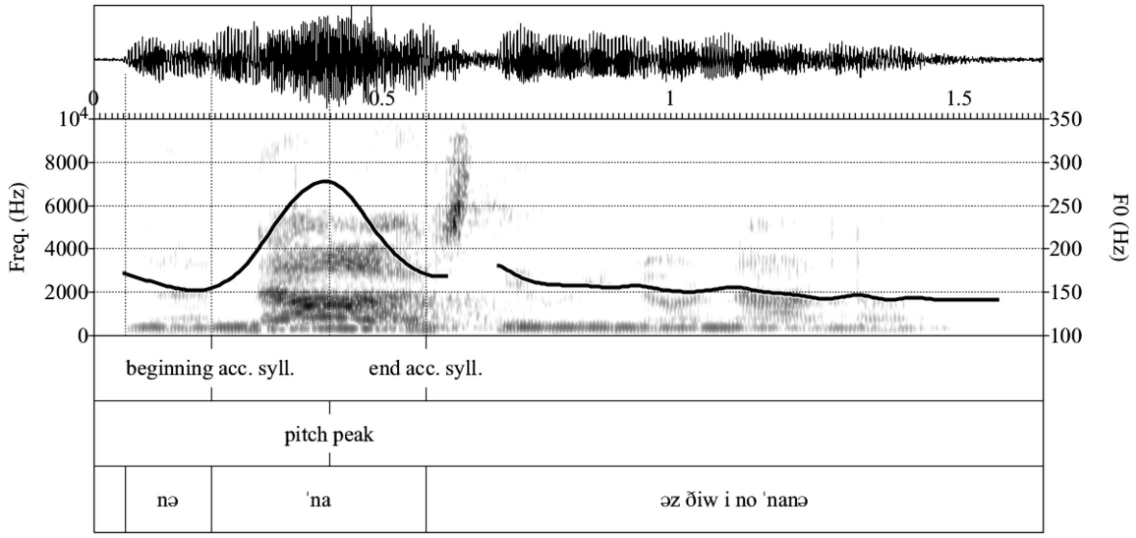




As for the acoustic analysis, three measures were annotated using the Praat software package (Boersma & Weenink, 2012) and then imported into ELAN: the beginning and end of the target focused word (tier 4 in Figure 5), the beginning and end of the accented syllable (tier 5 in Figure 5), and the pitch peak of the F0 line (tier 6 in Figure 5). Figure 6 is an example of the acoustic labeling in Praat that was later imported into ELAN. As it can be observed, Praat allows annotating the acoustic signal with a time set of 0.001 seconds. The precise milliseconds in which these acoustic measures occurred in Praat were later annotated in ELAN by means of the option “accessing points in time” that this annotation software offers. The pitch peak was easily located in the F0 line because target words were pronounced in a contrastive focus condition, so the rising contour was very salient. Following Knight & Nolan (2006), if participants did not pronounce it emphatically and it was as a consequence not so salient in the F0 line, the pitch peak was located at the end of the *plateau*. In this sense, Knight & Nolan (2006) showed that whenever a H (high) tone of a nuclear accent is realized as a flat stretch of contour rather than as a single turning point, the end of the plateau is the point that is anchored within the syllable and seems to be equivalent to the alignment of H turning points, being a marker of linguistic structure.

.

Figure 6. Example of acoustic labeling in Praat of the target word “na’na”



3. RESULTS

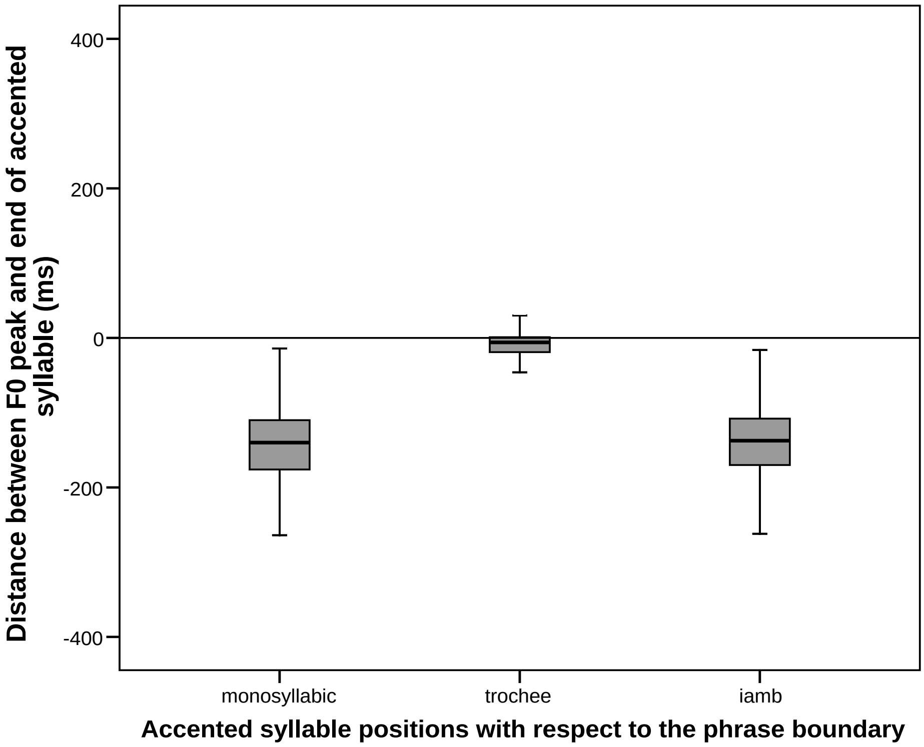
The results of the acoustic and gestural analysis will show the way gesture and prosodic structures are coordinated. The following research questions had to be answered: (1) Does the position of the intonation peaks within the accented syllable change when the position of the accented syllable with respect to the phrase boundary is different? If so, our results would corroborate previous findings in Silverman & Pierrehumbert (1990), Prieto et al. (1995), and D'Imperio (2001) in the sense that the position of the F0 peak varies depending on the position of the stress with respect to the phrase boundary; (2) In what way do the position of the apexes change with respect to the end of the accented syllable when the position of the accented syllable within the intonational phrase is different?; (3) Does the timing of intonation peaks correlate with the timing of gesture apexes? If results of question (1) and (2) show that intonation peaks and apexes vary depending on the distance of the accented syllable with respect to the boundary tone, we would show that intonation peaks and gesture apexes are synchronized and follow the same timing patterns.

Our results are presented in three subsections, one for each research question. The first subsection presents the results on the position of intonation peaks with respect to the accented syllable, taking the end of the accented syllable as the reference point to measure its position. The second subsection shows the results on the position of gesture apexes with respect to the end of the accented syllable, as well as results on the timing of the gesture onset with respect to the speech signal. And the third subsection presents the results obtained from the analysis of the third research question, namely whether intonation peaks and gesture apexes align with each other.

3.1 Position of the intonation peaks with respect to the end of the accented syllable

The data was analyzed using a repeated measures ANOVA (RM ANOVA). The independent factor was pitch accent position within the word (3 levels: monosyllables, trochees, iambs), and the dependent factor was the distance in milliseconds between the intonation peak and the end of the accented syllable[i]. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 13.111$, $p < .01$), and degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity (ε = 0.95). The statistical analysis revealed that the position of the intonation peaks within the accented syllable was significantly different depending on the pitch accent position ($F(1.894, 433.760) = 580.318$, $p < .001$, $\eta p^2 = .717$). As expected, Bonferroni pair-wise comparisons showed that the distance in time between intonation peaks and end of accented syllables did not vary significantly in monosyllables and iambs ($p = 1.000$), while this temporal distance was significantly different when comparing monosyllables and trochees ($p < .001$) and when comparing trochees and iambs ($p < .001$). Figure 7 shows the temporal distance (in ms) between intonation peaks and the end of the accented syllable as a function of the distinct pitch accent positions analyzed.

Figure 7. Box plots of the temporal distance (in milliseconds) between intonation peak and the end of the accented syllable as a function of the three distinct accented syllable positions within the word.

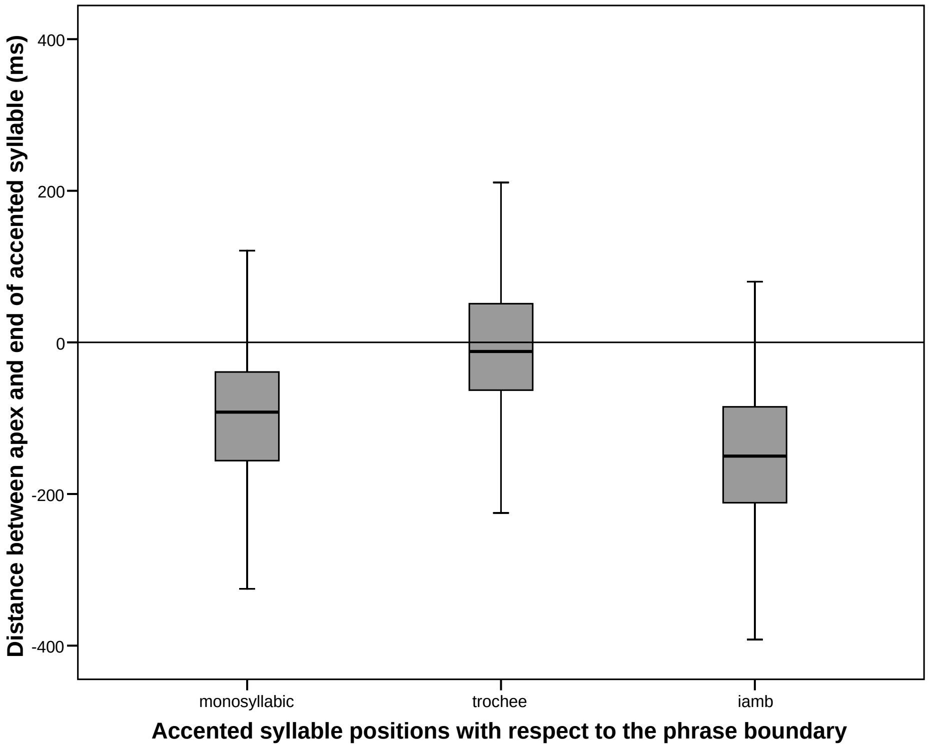


These results confirm previous findings (D'Imperio, 2001; Prieto et al., 1995; Silverman & Pierrehumbert, 1990) showing that the position of H intonation peaks varies depending on the distance to the upcoming boundary tone. Thus, intonation peaks are retracted in monosyllables and in iambs in a contrastive focus condition in such a way that the rising accent is realized as a complex rise-fall movement if in phrase-final position. Thus our results corroborate previous findings in the literature by showing that the intonation peaks are retracted in the case of monosyllables (a mean of 141.30 ms, *SD*=53.04) and iambs (a mean of 137.52 ms, *SD*=50.36) and that they are not retracted in the case of trochees (a mean of 11.90 ms, *SD*=39.1), thus occurring almost at the end of the accented syllable.

3.2 Position of the gesture apexes with respect to the end of the accented syllable

Again, the data was analyzed using a RM ANOVA. The independent factor was pitch accent position within the word (3 levels: monosyllables, trochees, iambs), and the dependent factor was the temporal distance between the apex and the end of the accented syllable (in milliseconds). Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 9.519$, $p < .01$), so the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.96$). The statistical analysis revealed that the position of the gesture apexes within the accented syllable was significantly different depending on the pitch accent position it occupies ($F(1.921, 440.008) = 196.675$, $p < .001$, $\eta p^2 = .462$). Bonferroni pair-wise comparisons showed that the distance between gesture apexes and the end of accented syllables varied significantly between monosyllables and iambs ($p < .001$), between iambs and trochees ($p < .001$), and also between monosyllables and trochees ($p < .001$). Figure 8 shows the temporal distance between the gesture apex and the end of the accented syllable as a function of the distinct pitch accent positions analyzed.

Figure 8. Box plots of the temporal distance (in milliseconds) between the apex and end of the accented syllable as a function of the three different positions of the accented syllable within the phrase.



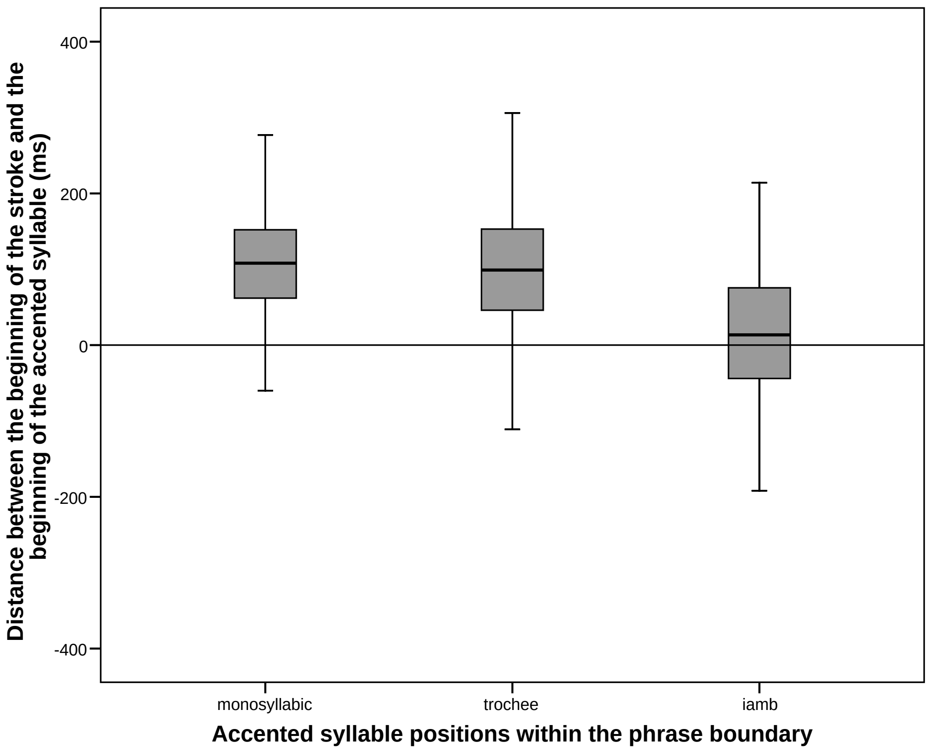
When the box plots of the distance between intonation peaks and the end of the accented syllable (Figure 7) are compared with the box plots of the distance between apexes and the end of the accented syllable (Figure 8), the same tendency can be observed. Both intonation peaks and apexes occur at the end of the accented syllable in trochaic words, that is, when the stress is in non-phrase-final position. And when the stress is in phrase-final position, intonation peaks and apexes occur well before the end of the accented syllable. However, the statistical analyses show that there is in fact a difference across the results of these two analyses. Specifically, monosyllables and iambs behave similarly in terms of the position of the intonation peak within the accented syllable, whereas they behave significantly differently in terms of the position of the gesture apex within the accented syllable, being less retracted in monosyllables than in iambs.

In order to find out the reason why gesture apexes are less retracted in monosyllables than in iambs whereas intonation peaks are equally retracted in both metrical structures, a closer analysis of the coordination between the gesture phases and the metrical structures was carried out. The specific metrical structure of monosyllabic words (i.e. words with the stress in phrase-final position and without pretonic segmental material) in comparison with the metrical structure of iambic words (i.e. words with the stress in phrase-final position and with pretonic segmental material) might be influencing the gesture phases and thus causing this ‘lagging effect’ of gesture apexes.

We hypothesize that the gesture onset starts being realized before with respect to the accented syllable when the stress is in phrase-medial position, and so the gesture apex is not “lagged” by a preceding phrase boundary. Then, this gesture onset timing makes the gesture stroke occur at different positions with respect to the onset of the accented syllable. Our hypothesis is that the specific metrical structure of monosyllabic words (i.e. words with the stress in phrase-final position and also adjacent to a phrase-initial boundary, without pretonic segmental material) in comparison with the metrical structure of iambic words (i.e. words with the stress in phrase-final position and with pretonic segmental material) might have had an influence on the realization of the timing of the preparation phase, thus causing a ‘lagging effect’ of gesture apexes. In order to test this hypothesis, two distinct analyses were carried out: first, to test whether the pretonic syllable in iambs contain part of the gesture preparation phase; second, to test whether this difference in timing was due to a difference in the gesture velocity or due to the different initiation time of the gesture with respect to the accented syllable.

First, in order to test if the pretonic syllable in iambs contains part of the gesture preparation phase, a RM ANOVA was carried out with the pitch accent positions as independent variable (3 levels: monosyllables, trochees, iambs) and the distance between the stroke onset and onset of accented syllable as dependent variable. Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2(2)$ = 2.559, *p* = .278), so sphericity was assumed (ε = 0.99). The statistical analysis revealed that the distance between stroke onset and onset the accented syllable was significantly different depending on the position of the pitch accent position in the target word (*F*(2, 458) = 108.203, *p* < .001, ηp^2 = .321). Bonferroni pair-wise comparisons showed that the distance between beginning of the stroke and beginning of the accented syllable varies significantly between monosyllables and iambs (*p* < .001), between trochees and iambs (*p* < .001), but not between monosyllables and trochees (*p* = 1.000). Figure 9 illustrates the mean distance in milliseconds between the beginning of the stroke and the beginning of the accented syllable as a function of the distinct pitch accent positions analyzed, showing that the stroke of the gesture starts when the accented syllable has already been initiated in monosyllables and trochees. By contrast, the stroke and accented syllable start almost simultaneously in the iambic condition. These results show that when the stress is in non-phrase-initial position, the pretonic syllable already contains part of the preparation phase of the gesture movement.

Figure 9. Box plots of the distance between the beginning of the stroke and the beginning of the accented syllable as a function of the three different accented syllable positions with respect to the boundary tone.

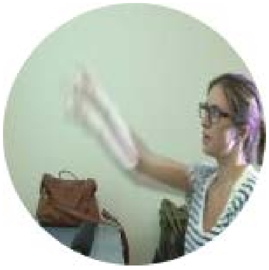
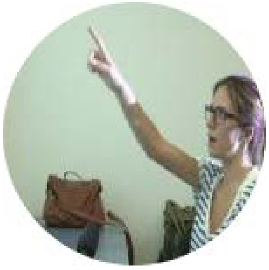
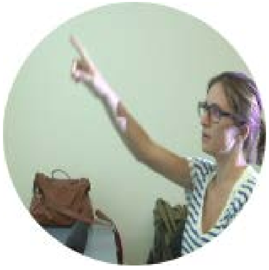
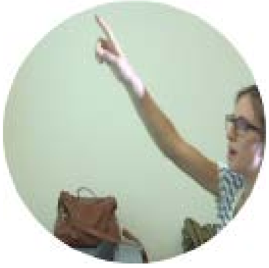
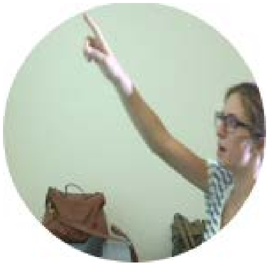
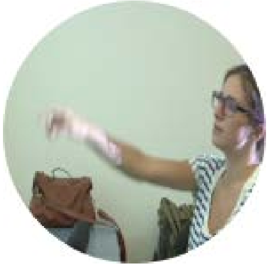


Furthermore, we run an additional RM ANOVA to test whether there the onset of the gesture was different across conditions. The distance between gesture onset and onset of accented syllable was the dependent variable, and stress position within the phrase was the independent variable. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2)$ = 187.376, *p* = .001), and degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity (ε = 0.637). The statistical analysis revealed that the distance between gesture onset and onset of accented syllable is significantly affected by the position of the accented syllable with respect to the phrase boundary (F(2, 284.071)=28.943, *p* < 001). Bonferroni post-hoc tests showed that monosyllables and trochees do not differ significantly (*p* = 1.000) whereas iambs differ with respect to monosyllables (*p* < .001) and trochees (*p* < .001). The mean values of this distance reveal that in monosyllables the gesture starts 460.5 ms before the accented syllable, in iambs the initiation of the gesture occurs 576.38 ms before the accented syllable starts being pronounced, and in trochees it occurs 475.82 ms the accented syllable. Thus, the gesture is initiated at different points in time with respect to the speech signal, being initiated before in iambs than in the other two conditions.

Second, in order to test if this difference in the timing of peak and speech prominence was due to a difference in the duration of the preparation phase (or launching of the pointing gesture, a measure which is indirectly related to the velocity of the launching movement), an additional RM ANOVA was applied. The distance between gesture onset and gesture apex was the dependent variable, and pitch accent position was the independent variable. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 172.632$, $p < .001$), and degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.649$).The statistical analysis revealed that the distance between these two points is not significantly affected by the pitch accent position ($F(2, 289.516)=2.699$, $p = .091$). This result suggests that since the time of the launching gesture does not differ across conditions, velocity profiles of the launching times do not seem to be responsible for the different location of the gesture apexes.

All together, the results of the previous analyses offer an explanation for why the gesture apexes occur later in monosyllables than in iambs with respect to the accented syllable: the gesture has to start later in monosyllables than in iambs (because they do not have pretonic material to contain the preparation phase), and so the apex is also "pressed" to the end of the accented syllable in monosyllabic condition due to a lagging effect. Table 1 shows several still images of the timing targets of the pointing in the two conditions where the stress is in phrase-final position. At the beginning of the accented syllable the arm is realizing the preparation phase in monosyllables (top left) whereas it is already at the stroke phase in iambs (bottom left). Then, when the accented syllable ends, the arm is still extended in monosyllables (top right), whereas in iambs it is already in the retraction phase (down right). These findings suggest that prosodic phrase boundaries strongly determine the timing of the gesture phases and can cause retracting effects (when there is an upcoming phrase boundary) as well as lagging effects (when there is a preceding phrase boundary).

Table 1. Still images of arm positions in three specific moments of the speech signal (the beginning of the accented syllable, the F0 peak, and the end of the accented syllable) in the monosyllabic condition and in the iambic condition.

	Beginning of accented syllable	**F0 peak**	**End of the accented syllable**
Monosyllables			
Iambs			

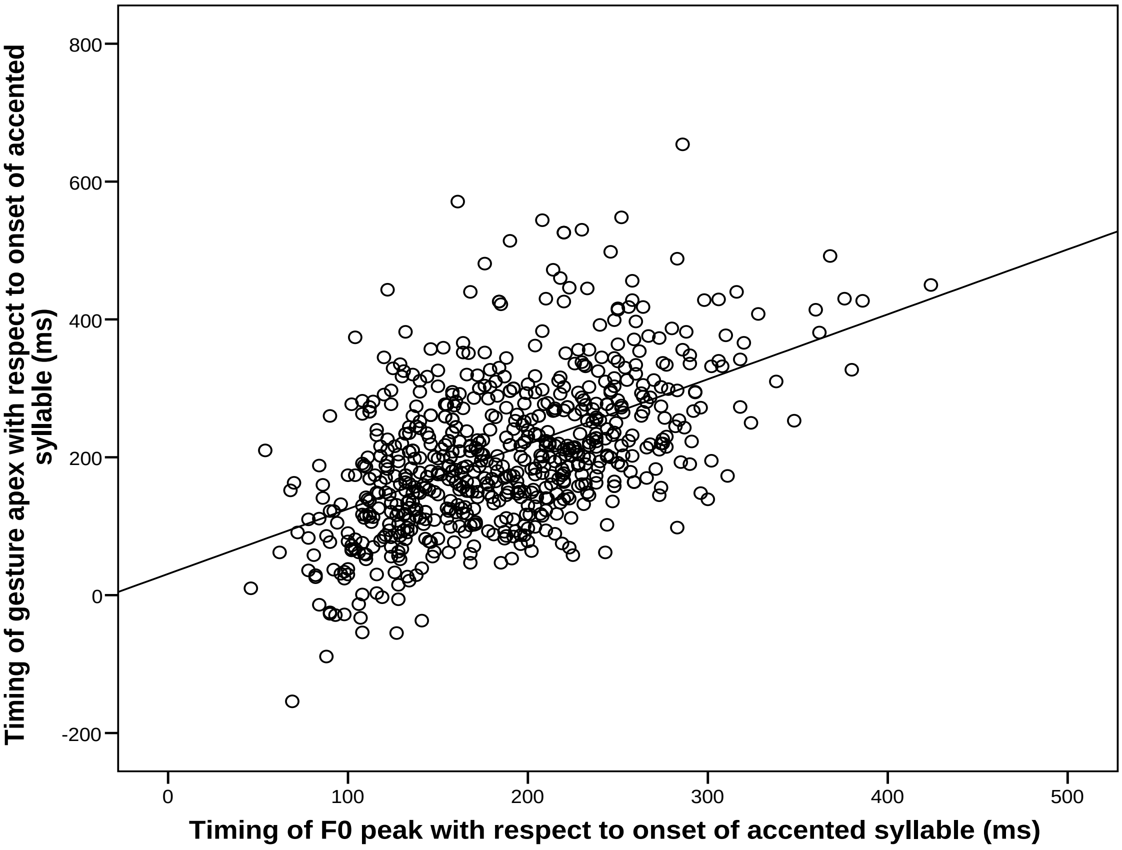
3.3 Does the position of intonation peaks correlate with the position of gesture apexes?

The goal of our study was to investigate which anchoring regions in speech and in gesture align with each other, hypothesizing that it is the intonation peak and the gesture apex that are temporally coordinated. A Pearson correlation analysis was carried out with two independent variables: timing of the gesture apex with respect to the onset of the accented syllable and timing of the F0 peak with respect to the onset of the accented syllable. Results showed that the timing of gesture apexes and F0 peaks with respect to the onset of the accented syllable was significantly correlated (*r*(703) = .528, *p* < .01). Figure 10 illustrates that the timing of gesture apexes and the timing of the intonation peaks is correlated.

Interesting results are obtained when comparing the correlation between gesture apex and F0 peaks (revealed as significant in our study) with other values that previous literature have suggested to play a crucial role in the temporal coordination of gesture and speech prominence: the end of the accented syllable (as an alternative to F0 peak), and the stroke onset and stroke offset (as alternatives to the gesture apex).

Thus, three additional Pearson correlation analyses were carried out. In the first correlation, the two independent variables were the gesture apex and the end of the accented syllable (as opposed to the F0 peak) with respect to the onset of the accented syllable. The analysis revealed that even though the timing of the gesture apex and the end of the accented syllable were significantly correlated, its low correlation coefficient value, or *r* values indicates that they are weakly correlated . (*r*(703) = .276, *p* = .01). In the second additional correlation, the two independent variables were the end of the stroke (as opposed to the gesture apex) and the F0 peak with respect to the onset of the accented syllable. Results show that these two variables are also significantly correlated, yet to a lesser extent than in the apex-intonation peak analysis (*r*(703) = .451, *p* = .01). In the third additional correlation, the two independent variables were the stroke onset (as opposed to the gesture apex) and the F0 peak with respect to the onset of the accented syllable. Results show that these two variables were significantly (but weakly) correlated (*r*(703) = .397, *p* = .01). All together, these correlation analyses reveal that given the *r* correlation coefficient values, the two measurements which are more strongly correlated are the ones we predicted: the timing of the gesture apex with the timing of the F0 peak.

Figure 10. Scatter plot showing the relation between the timing of the gesture apexes and F0 peaks with respect to the onset of the accented syllable.



4. DISCUSSION

This study had two main aims, namely (a) to determine the anchoring region in speech that aligns with the prominent part of the gesture in a contrastive focus condition; and (b) to test whether intonation and gestural movements behaved in a similar way and they are both bound by prosodic structure. Our hypotheses were (a) that it is the peak of the focused pitch accent in the intonational structure that serves as an anchoring site for gestural prominence; and (b) that upcoming L boundary tones affect the timing of both intonation and gesture movements. In order to tease out both research questions, this study analyzed 703 instances of co-speech pointing deictic gestures produced by monolingual adult Catalan speakers. During the experiment, the speech consisted of the pronunciation of a target word followed by a boundary tone in a contrastive focus condition. When designing the experiment, it was crucial to make sure that participants produced target focused words with different metrical patterns in a contrastive focus condition, which is typically realized prosodically with a prominent L+H* pitch accent and followed by a L- boundary tone. This automatically triggered distinct positions of the intonation peaks, which were retracted in the case of stress-final words.

After analyzing our data, two main findings were obtained: first, that the timing of gesture apexes behaves very much like intonation peaks with respect to their position within the accented syllable, that is, both were retracted when a prosodic phrase was present at the end of the accented syllable ; and second, that the position of the peak of the pitch movement (intonation peak) and the position of the peak of the gesture movement (apex) were the two measurements with a stronger correlation. We now discuss these two main findings.

The first main finding was that the temporal behavior of intonation peaks and apexes is similar. On the one hand, the statistical analyses showed that the distance between intonation peaks and the end of the accented syllable varied significantly depending on the position of the pitch accent in the target word with respect to the upcoming phrase boundary: when the target word had the stress in non-phrase-final position, intonation peaks were produced closely aligned with the end of the accented syllable. However, when the target word had the stress in phrase-final position, i.e. in monosyllables and iambs, intonation peaks were produced well before the end of the accented syllable. On the other hand, the statistical analyses showed that the position of gesture apexes within the accented syllable also varied significantly depending on the position of the pitch accent with respect to the phrase boundary: the apex was produced closely aligned with the end of the accented syllable when the accented syllable was in non-phrase-final position, and it was produced well before the end of the accented syllable when the accented syllable was in phrase-final position. In conclusion, time-wise both intonation peaks and gesture apexes occur at the end of the accented syllable when the stress is not in phrase-final position, while they are temporally retracted when the stress is in phrase-final position.

The temporal retraction of intonation peaks within the accented syllable before upcoming prosodic boundaries is a well documented phenomena in Catalan as well as in other languages (D'Imperio, 2001; Prieto et al., 1995; Silverman & Pierrehumbert, 1990). It is caused by the fact that when H* accents occur in phrase-final nuclear position before L% boundary tones, they are moved to the left by the adjacency of the boundary tone, thus retracting the intonation peak towards the beginning of the accented syllable. By contrast, in non-phrase final nuclear position, the accented syllable does not have to accommodate the entire complex pitch movement: the rising pitch movement is realized during the accented syllable, and the falling pitch movement is realized during the unaccented syllable. Similarly, the same 'temporal adaptation' phenomenon might be taking place with gestures. The temporal behavior of gesture apexes in our experiment shows that when the nuclear pitch accent is in phrase-final position, the gesture apex needs to retract for the arm to have time to return at its rest position within the limits of the target segmental material,. In a non-phrase final condition, however, since the arm's return to its rest position does not have to be completed by the end of the accented syllable (which is not in phrase-final position) and thus the rest of the arm trajectory can be accommodated during the posttonic unaccented syllable.

Interesting results were obtained when analyzing the gestural timing patterns when the focal pitch accent was in phrase-final position, namely that the gesture apex is anchored at a different position within the accented syllable when comparing monosyllables and iambs. The statistical analyses showed that gesture apexes are retracted in both monosyllables and iambs, but that the amount of retraction is significantly different. In iambs, the gesture apex occurred a mean of 141.11 ms before the end of the accented syllable, whereas in monosyllables the gesture apex occurred a mean of 89.12 ms before the end of the accented syllable (*see* Figure 7, where this difference becomes clear). In order to understand why gesture apexes were more retracted in iambs than in monosyllables, we analyzed in detail the timing of the arm launching movement with respect to the timing of speech in both metrical structures. These analyses showed that the distance to the beginning of the prosodic phrase had a direct influence on the timing of the gesture phases: when the stress is in phrase-initial position (in monosyllables and iambs), the preparation phase of the pointing gesture has to start during the accented syllable because there is no preceding unstressed syllables to anchor the preparation phase to. However, when the stress is in phrase-medial position (in iambs), the preparation phase can start before the accented syllable because there is a previous unaccented syllable that can accommodate part of the preparation phase. Further analyses showed that these results are due to a different initiation time of the gesture movement with respect to the accented syllable more than to velocity effects: the initiation of the gesture with respect to the onset of the accented syllable occurs well before in iambs (where there is an preceding unstressed anchoring syllable) than in monosyllables (where the accented syllable is in phrase-initial position).

The monosyllabic condition turned out to be a very interesting case because it was the only condition in which the target syllable was both adjacent to a phrase-initial and a phrase-final position: an adjacent phrase-final boundary caused a ‘retraction effect’ in the location of the gesture apex, but an immediately preceding start of a phrase made this retraction not so evident as in the iambic case, causing what we called a ‘lagging effect’. Thus, the timing of the launching gesture had to adjust to both utterance-initial and utterance-final prosodic requirements because there was no other segmental material between the phrase boundaries where to accommodate the gestural initiation and retraction phrases. By contrast, when there was a pretonic syllable available (as in iambs) or a posttonic syllable available (as in trochees), the lagging effect was not observable in the timing of the gesture peaks.

The second main finding was that the position of the peak of the pitch movement (intonation peak) and the position of the peak of the gesture movement (apex) were the two measurements in gesture and speech with a higher correlation, thus corroborating our initial hypothesis. Some previous studies suggested that the prominent part of the gesture coordinates with the stressed syllable (Loehr, 2004, 2007; Rochet-Capellan et al., 2008), while others found that only stressed syllables containing an intonation peak (De Ruiter, 1998; Nobe, 1996) coordinate with the prominence in gesture. In our study we defined more precisely the assumption that the prominent part of the gesture (the stroke phase) coordinates with the speech prominence (the accented syllable): the location of the gesture apex with respect of the accented syllable is highly correlated with the location of the intonation peak within the accented syllable, in a contrastive focus condition and when followed by a phrase boundary. Comparing our results on the correlation between intonation and gesture peaks with other measurements taken into account in the literature (namely, the accented syllable in the speech modality and the stroke onset or stroke offset in the gesture modality), our results crucially show that the apexs time location is less correlated with the end of the accented syllable than with the F0 peak, and that F0 peaks are less correlated with stroke onset or stroke offset than with gesture apexes, as gesture apexes and F0 peaks are both bound by prosodic structure.

All in all, our results show that prosodic structure influences the timing of gesture movements in two main ways: first, the gesture apex always occurs during the accented syllable, irrespectively of the position of the accented syllable with respect to the phrase boundary; and second, the position of the gesture apex is highly correlated with the position of the F0 peak, irrespectively of their position within the accented syllable triggered by an upcoming phrase boundary tone. These results are in line with Rochet-Capellan et al.'s (2008) study in the sense that the position of the accented syllable within the phrase affected the timing of gesture phases. Similarly, our results confirm De Ruiter's (1998) second experiment in which the stress location directly affected the timing of the gesture: in this data, the apex of the gesture occurred during the accented syllable, regardless of the metrical structure of the target word.

In addition, our findings confirm that the gesture apex is an important measure to take into account in when studying temporal coordination between deictic gestures and speech (in line with De Ruiter 1998, McClave 1994, Rochet-Capellan et al. 2008, Roustan & Dohen 2010, and Rusiewicz 2010). When trying to characterize gesture-speech prominence alignment, other gesture measurements taken into account in the literature, such as the gesture onset or the gesture stroke do not seem to be so telling. In our results, it is precisely the gesture apex that is strongly correlated with the intonation peak, independently of the position of the accented syllable within the prosodic phrase.

The main contribution of this article is to show that both intonation and gesture movements are bound by prosodic phrasing, causing retracting effects when there is an upcoming phrase boundary (as in monosyllables and iambs), and also lagging effects when there is no pretonic or posttonic syllable after a preceding phrase boundary to contain part of the gesture prominence (in monosyllables). We also defined in a precise way the temporal coordination between gesture and speech in a narrow contrastive focus condition, showing that the anchoring point for the gesture apex is the intonation peak. By analyzing deictic gestures accompanied by target words with different metrical structures followed by a phrase boundary and in a contrastive focus condition, we showed that gesture apexes align with intonation peaks, and that both intonation and gesture structures have a parallel temporal behavior: when F0 peaks are produced at the end of the accented syllable, gesture apexes are also produced at the end of the accented syllable (e.g. in the trochaic condition); and when F0 peaks are retracted within the accented syllable, gesture apexes are also retracted within the accented syllable (e.g. in the iambic condition). Gesture apexes and F0 peaks are closely aligned even in monosyllabic words, whose gesture apexes are influenced by a ‘lagging effect’ (because of the presence of a preceding prosodic phrase boundary) and by a ‘retracting effect’ (because of an upcoming prosodic phrase boundary). Hence, the results of our study suggest that gesture structure adapts to prosodic structure and that pointing gesture movements are bound by prosodic phrasing the same way that pitch movements are.

5. CONCLUSION

Our study confirms the results of most of the previous studies investigating the temporal coordination between gesture and speech, namely that the prominence in gesture and prominence in speech are temporally coordinated. Previous findings in the literature have showed that the most meaningful part of the gesture (i.e. the stroke or the apex) was temporally synchronized either with the focused word (Butterworth & Beattie, 1978; Roustan & Dohen, 2010), or with the lexically stressed syllable (Loehr, 2004, 2007; Rochet-Capellan et al., 2008), or with syllables with intonation peaks (Nobe, 1996; De Ruiter, 1998 –*second experiment*).

After analyzing pointing gestures accompanied by target words with varying positions of the accented syllable with respect to the phrase boundary, our results showed not only that the most prominent part of the gesture (the apex) is coordinated with the accented syllable, but also that the anchoring region in speech that synchronizes with the gestural prominence is the intonation peak. This study shows that, in a contrastive focus condition, the specific point in the speech signal that is correlated with the apex of deictic gestures is the intonation peak rather than the interval of time where the accented syllable is produced. Also, our findings suggest that both intonation and gesture movements are bound by prosodic phrasing because the timing of their starting movements and prominence peaks (F0 peak and apex) varies if there is a preceding or an upcoming prosodic phrase boundary.

Further research is needed until we have a complete description of how gesture and speech are temporally synchronized. First, our hypothesis should also be tested in a natural environment rather than in an experimental setting. The challenge is to obtain spontaneous instances of spontaneous co-speech deictic gestures in different types of focus conditions. Second, these precise analyses could also be extended to other types of gestures in order to investigate the specific anchoring regions in prosodic and gesture structures that are temporally coordinated. Third, it will be interesting to test the alignment of gesture apexes with F0 peaks (and F0 falls) further, with experiments in which the F0 did not occur within the accented syllable, or in with languages with distinct alignment patterns of rising and falling pitch accents within the accented syllable. Finally, future experimental studies will have to analyze (a) the role of temporal coordination patterns between gesture and prosody for successful language communication, and (b) the role of temporal coordination patterns for language comprehension. Our hypothesis is that the lack of precise coordination between gesture and speech affect the perception, comprehension, and integration of the gesture-speech combination in the brain. If this is indeed the case, we can be confident that our results are important for a full understanding of human communication.

6. ACKNOWLEDGMENTS

A preliminary version of this paper was presented at the conferences *Architectures and Mechanisms for Language Processing* (Paris, France, September 1-3, 2011), *Gesture and Speech in Interaction* (Bielefeld, Germany, September 5-7, 2011), *V International Conference for Gesture Studies* (Lund, Sweden, July 24-27, 2012), and *Laboratory*

Phonology (Stuttgart, Germany, July 27-29, 2012). We would like to thank participants at those meetings, especially Mark Swerts, Heather Rusiewicz, Stefan Kopp, Zofia Malisz, Stefanie Shattuck-Hufnagel, Adam Kendon, John J. Ohala, and Robert Port for their helpful comments. We thank Aurora Bel, Louise McNally, and José Ignacio Hualde for being part of the PhD project defense and particularly for their useful advice regarding the methodology used in this study. We also thank Joan Borràs for his help with running the experiment and carrying out the statistical analysis, and Paolo Roseano for his comments on Praat annotations.

This research has been funded by two research grants awarded by the Spanish Ministerio de Educación y Ciencia, FFI2009-07648/FILO “The role of tonal scaling and tonal alignment in distinguishing intonational categories in Catalan and Spanish” and by a grant awarded by the Generalitat de Catalunya to the *Grup d’Estudis de Prosòdia* (2009SGR-701).

7. **REFRENCES**

Birdwhistell, R. L. (1952). *Introduction to Kinesics: An Annotated System for Analysis of Body Motion and Gesture*. Washington D.C.; Dept. of State, Foreign Service Institute.

Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body Motion Communication.* Filadelfia: University of Pennsylvania Press.

Boersma, P. & Weenink, D. (2012). *Praat: doing phonetics by computer* [Computer program]. Version 5.3.04, retrieved 12 January 2012 from http://www.praat.org/.

Butterworth, B. & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. Campbell & G.T. Smith (Eds.), *Recent advances in the psychology of language: formal and experimental approaches*, 347–360. New York: Plenum Press.

Butterworth, B. & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, *96*(1), 168–174.

D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication 33*(4), 339-256.

De Ruiter, J. P. (1998). *Gesture and speech production*. Unpublished doctoral dissertation, Katholieke Universiteit, Nijmegen, Germany.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics, 39*(10), 1722-1741.

Kendon, A. (1972). Some relations between body motion and speech. An analysis of an example. In Siegman, W. & Pope, B. (Eds.). *Studies in Dyadic Communication*, 177-210. New York: Pergamon Press.

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*, 207-227. The Hague: Mouton.

Knight, R. A. & Nolan, F. (2006). "The effect of pitch span on intonational plateaux". *Journal of the International Phonetic Association*, *36*(1), 21–38.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, *61*(5), 743–754.

Lausberg, H. & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, *41*(3), 841–849.

Levelt, W. J., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*, 133–164.

Loehr, D.P. (2004). *Gesture and Intonation*. Unpublished doctoral dissertation, Georgetown University.

Loehr, D.P. (2007). Aspects of rhythm in gesture and speech. *Gesture*, *7*, 179–214.

McClave, E. (1994). Gestural beats: the rhythm hypothesis. *Journal of Psycholinguistic Research*, *23*, 45–66.

McNeill, D. (1992). *Hand and Mind*. Chicago, London: The Chicago University Press.

Nobe, S. (1996). *Representational Gestures, Cognitive Rhythms, and Acoustic Aspects of Speech: A Network/Threshold Model of Gesture Production*. Unpublished doctoral dissertation, University of Chicago.

Prieto, P. (1995). Aproximació als contorns entonatius del català central. *Caplletra, 19*, 161–186.

Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal Alignment Patterns in Spanish. *Journal of Phonetics, 23*, 429–451.

Prieto, P. & Ortega-Llebaria, M. (2009). Do complex pitch gestures induce syllable lengthening in Catalan and Spanish?. In M. Vigário, S. Frota & M. J. Freitas (Eds.). *Phonetics and Phonology: Interactions and Interrelations*, 51–70. Amsterdam/Philadelphia: John Benjamins.

Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. L. (2008). The speech focus position effect on jaw-finger coordination in a pointing task. *Journal of Speech and Language Hearing Research 51*(6), 1507–1521.

Roustan, B. & Dohen, M. (2010). Co-Production of Contrastive Prosodic Focus and Manual Gestures: Temporal Coordination and Effects on the Acoustic and Articulatory Correlates of Focus. *Proceeding of the Speech Prosody 2010*. Chicago, IL, USA.

Rusiewicz, H. L. (2010). *The role of prosodic stress and speech perturbation on the temporal synchronization of speech and deictic gestures*. Unpublished doctoral dissertation, University of Pittsburgh.

Silverman, K. & Pierrehumbet, J. (1990). The timing of prenuclear high accents in English. In Kingston, J. & Beckman, M. (Eds.) *Papers in Laboratory Phonology I*, 72–106. Cambridge: Cambridge University Press.

Vanrell, M. M. (2011)."The phonological relevance of tonal scaling in the intonational grammar of Catalan". PhD Dissertation. Bellaterra: Universitat Autònoma de Barcelona.

[i] In order to test if the consonant status had an effect on the timing of intonation peak, a RM ANOVA had been carried out with consonant voicing as independent factor (2 levels: voiced nasals, unvoiced stops), and distance in ms. between intonation peak and end of accented syllable as the dependent factor. The analysis indicated that the position of the F0 peak within the accented syllable was not significantly affected by the voicing of the consonant ($F(1,344) = .012$, $p = .913$). This result confirms previous claims on the location of the F0 H peaks for rising pitch accents in contrastive focus in Catalan: F0 peak are consistently located at the end of the accented syllable, regardless of variation in consonant identity (e.g., Vanrell 2011).