The timing of head nods is constrained by prosodic structure

Escrit per: Eli Asor

Dirigit per: Prof. Pilar Prieto i Prof. Joan Borràs-Comes

Barcelona, 11 d'agost de 2014
ABSTRACT

Purpose: previous studies in the field of prosody-gesture alignment have found that prominent parts of gestures are temporally aligned with prosodically prominent parts of the utterance. Yet, little is known about the precise alignment of gestures with respect to the prosodic structure of the associated speech. Hypotheses: in confirmation utterances (a) the position of the nuclear accented syllable with respect to an upcoming prosodic boundary influences the timing of the intensity peak within that syllable, (b) the position of the nuclear accented syllable with respect to an upcoming prosodic boundary influences the timing of the associated head nod gesture, and (c) the apex of the head nod gesture is temporally aligned with the intensity peak of the nuclear accented syllable. Method: 11 Catalan speakers performed a Discourse Completion Task involving confirmatory contexts. Recordings of spontaneously occurring co-speech head nod gestures were acoustically and gesturally analyzed. Results: head nod apexes and intensity peaks are temporally associated with the end of the accented syllable in words with penultimate and antepenultimate stress, while they are temporally retracted in monosyllables and iambs when immediately preceding a prosodic boundary. Conclusions: these results expand the findings by Esteve-Gibert & Prieto (2013), showing that the timing of gesture apexes in head nod gestures produced in semi-spontaneous speech is conditioned by the proximity to prosodic boundaries, as well as by the metrical structure of the prominent word in the utterance. Moreover, this study shows that in addition to the intonational peak of the nuclear accented syllable, its intensity peak also highly correlates with the gesture apex.
# CONTENTS

1. Introduction .......................................................................................................................... 4

2. Methodology .......................................................................................................................... 7
   2.1. Participants ....................................................................................................................... 7
   2.2. Materials ......................................................................................................................... 7
   2.3. Procedure ....................................................................................................................... 8
   2.4. Coding ............................................................................................................................ 9

3. Results ..................................................................................................................................... 13
   3.1. Timing of the intensity peak ........................................................................................... 14
   3.2. Timing of the head nod apex .......................................................................................... 16
   3.3. Intensity-gesture synchrony ............................................................................................ 17

4. Discussion ............................................................................................................................. 19

References ................................................................................................................................... 22

Appendix ...................................................................................................................................... 25
1. **Introduction**

In recent years, many studies dealing with the gesture-speech interface in human communication have shown that these two components are highly integrated. In an earlier article, McNeill (1985: 353) suggested that gestures and speech form part of a single communication system, or in his own words, that they "are part of the same psychological structure and share a computational stage". He motivated this proposal using five arguments: (a) gestures occur only during speech, (b) they have pragmatic and semantic functions parallel to those of speech, (c) they are synchronized with linguistic units in speech, (d) they dissolve together with speech in aphasia, and (e) they develop together with speech in children. Based on these arguments, McNeill (1992) proposed three "synchrony rules", of which the "phonological synchrony rule", based on argument (c) above, is of the highest relevance to the present work.

McNeill based his formulation of the "phonological synchrony rule" on Kendon's (1980) words to state that "the stroke of the gesture precedes or ends at, but does not follow, the phonological peak syllable of speech". Nobe (1996) refined the "phonological synchrony rule" and proposed the "gesture and acoustic-peak synchrony rule". He defined the acoustic peak as a peak of either the fundamental frequency or the intensity (or both if they co-occur) and stated that the gestural stroke precedes or co-occurs with, but does not follow, the later of the two types of acoustic peak (if they do not co-occur). Contrary to those findings, McClave's (1994) study on beat gestures\(^1\) found that not all gestural beats co-occurred with stressed syllables. What she found was that gestural beats seemed to be generated rhythmically from the tone unit nucleus\(^2\) outwards, in such a way that they are found at even intervals from the nucleus, even if they fall on unstressed syllables or pauses.

Setting beat gestures aside, other studies have also refined the description of the temporal synchrony of gesture and speech. De Ruiter (1998), conducted two production experiments in which Dutch speakers had to point at a picture while naming the object it

\(^1\) Following McNeill’s (1992; p. 80) terminology, beat gestures are defined as "movements that do not present a discernible meaning, and they can be recognized positively in terms of their prototypical movement characteristics. They are typically biphasic (two movement components), small, low energy, rapid flicks of the fingers or hand (...)."

\(^2\) According to Cruttenden’s (1997; pp. 29-37) criteria.
showed. The first experiment tested new-information focus statements and the second one tested contrastive focus statements. While the author did not find a strong effect of metrical structure on the gesture apex onset, in any of the two conditions, he did find a robust effect of metrical structure on the gesture apex duration in the contrastive focus condition (i.e., the later the stressed syllable occurred, the longer the apex lasted). This led De Ruiter (1998) to formulate the "strict phonological synchrony rule" (i.e., the apex does not only precede the stressed syllable, but its timing also co-varies with the position of the stressed syllable). Loehr (2004, 2007) analyzed the different types of gestures produced by 15 native speakers of American English in spontaneous conversations and found, among other findings, that apexes of manual gesture strokes highly correlated with pitch accents, and that gesture phrases typically correlated with intermediate phrases of the co-occurring speech. A more recent study by Esteve-Gibert & Prieto (2013) investigated the precise alignment between deictic pointing gestures and fundamental frequency peaks in Central Catalan contrastive focus statements, controlling for three metrical structures (monosyllables, iambs and trochees). Their results showed that gesture apexes aligned with fundamental frequency peaks as well as a clear effect of metrical pattern with respect to upcoming prosodic boundaries on the temporal location of the gestural targets (apex, stroke and onset of the gesture).

While the temporal alignment of gesture and speech has been studied extensively in the field of manual gestures, there is less research on the way head gestures align with speech. Head gestures have been found to provide powerful cues for speech perception and comprehension when synchronized with the stressed vowel of the potentially prominent word in Swedish, especially in conveying prominence (House et al. 2001) and for word recognition (Al Moubayed & Beskow 2009). Prieto et al. (2011) found that, in

---

3 Rusiewicz (2010) also studied the influence of metrical structure on the alignment of deictic pointing gestures in contrastive focus conditions. She found that English native speakers’ gesture apexes were not aligned with the stressed syllable, but rather with the target word’s onset. In another experiment, she found that speech and gesture where less synchronous when produced in a delayed auditory feedback condition. However, she stated that changes in total gesture time, gesture launch time and gesture return time between the two conditions might suggest a potential interaction and feedback between the two systems.

4 He looked at iconic, metaphoric, deictic and beat manual gestures (following the classification by McNeill 1992: 75-82)

5 See Wagner, Malisz & Kopp (2014) for an extensive overview.
Catalan, visual cues were more powerful than acoustic ones for speakers’ ability to distinguish narrow from contrastive focus statements. In a second experiment, they found that, within the visual domain, head inclination served as a stronger perceptual cue than eyebrow raising did for this distinction. Sargin et al. (2005) analyzed semi-spontaneous speech of Canadian English speakers and found a high correlation between timings of accented syllables and start and end timings of head nod gestures. Alexanderson et al. (2013a, 2013b) found, in an experiment with Swedish speakers, that the point at which a head nod was in its lowest position was closely aligned with the nucleus of the accented syllable, while the point of maximum head velocity was closely aligned with the onset of the accented syllable. Finally, in a recent production study, Esteve-Gibert et al. (2014) analyzed the alignment of head gestures with co-occurring speech in a semi-spontaneous experimental setting, controlling for three types of metrical structures with respect to an upcoming phrase boundary (monosyllables, trochees and iambs). The alignment patterns they found were similar to those reported in Esteve-Gibert & Prieto (2013), in that gesture apexes were closely aligned with the end of the accented syllable in trochees, while they were temporally retracted in monosyllables and iambs.6

To the best of our knowledge, Esteve-Gibert et al. (2014) is the only study that accurately analyzes the patterns of temporal alignment of head gestures and speech, comparing two points in time (rather than intervals), with respect to prosodic structure. However, they did not attempt at comparing the temporal correlation of the gesture apex against different prominence points in speech, thus possibly missing the accurate anchoring point of the head gesture apex in speech.

The purpose of this study is, therefore, to address this issue using additional acoustic measures, as well as to contribute empirical data on the temporal coordination of head gestures, focusing on confirmatory head nods obtained in semi-spontaneous speech. Building on Esteve-Gibert et al. (2013, 2014) and on Nobe (1996), we hypothesize that in confirmation utterances (a) the position of the nuclear accented syllable with respect to an

---

6 Loehr (2004: 92-93) also looked at head gestures, in addition to the manual gestures his participants performed, but as he stated in the Methodology section of his study, he chose to exclude the head gestures from some of his analyses, since it was less clear to him what constituted the preparation, retraction, hold or resting position for the head.
upcoming prosodic boundary influences the timing of the intensity peak within that syllable, (b) the position of the nuclear accented syllable with respect to an upcoming prosodic boundary influences the timing of the associated head nod gesture, and (c) the apex of the head nod gesture is temporally aligned with the intensity peak of the nuclear accented syllable. In addition, to the best of our knowledge, this is also the first study to look at five different types of metrical structures (i.e., monosyllables, trochees, dactyls, iambs and amphibrachs) with respect to their temporal alignment with co-occurring spontaneous head gestures.

2. **Methodology**

2.1. **Participants**

The participants in this study were 11 Central Catalan speakers (4 male, 7 female), between 22 and 54 years of age (mean 30.5), all of them students or staff at the Universitat Pompeu Fabra in Barcelona. They participated voluntarily and were not aware of the purpose of the experiment.

2.2. **Materials**

The materials for this study consisted of a set of 25 cards, each containing a Discourse Completion Task (henceforth, DCT; Blum-Kulka et al. 1989, Billmeyer & Varghese 2000, Félix-Brasdefer 2010). Each DCT described a situation in which a hypothetical interlocutor hesitates whether or not a certain city (whose name appeared on the card) was the capital of a certain country, a Spanish autonomous community or a district in Catalonia (whose name also appeared on the card). The target words in each DCT had either of 5 different metrical structures, as described in Table 1. We chose to use names of capital cities as target words so that the situations described in the DCTs would be as close as possible to natural conversation situations. Five target words were chosen for each metrical structure, resulting in a total of 25 items per participant. These target items were embedded in either of 5 DCT templates, all of which consisted of social situations in which the participants

---

7 DCTs are widely used in the field of cross-cultural pragmatics to elicit speech acts. They contain a situational prompt which is either read to- or by the participant, in order to elicit his/her response.
were asked to situate themselves. The purpose of the different context templates was to make participants pay attention to the contexts presented to them.

<table>
<thead>
<tr>
<th>Syllabic Structure</th>
<th>Stress Position</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>monosyllabic</td>
<td>S</td>
<td>Vic, Valls</td>
</tr>
<tr>
<td>iambic</td>
<td>w-S</td>
<td>Paris, Dakar</td>
</tr>
<tr>
<td>trochaic</td>
<td>S-w</td>
<td>Roma, Lima</td>
</tr>
<tr>
<td>dactylic</td>
<td>S-w-w</td>
<td>Mónaco, Washington</td>
</tr>
<tr>
<td>amphibrach</td>
<td>w-S-w</td>
<td>Figueres, Caracas</td>
</tr>
</tbody>
</table>

Table 1: The different metrical structures controlled for in this study. 'S' represents a strong (stressed) syllable; 'w' represents a weak (unstressed) syllable. In the examples column, stressed syllables are underlined.

Example (1) demonstrates one of the DCTs used in this study\(^8\) (the target word in this example is Roma):

(1) Esteu jugant al Trivial i tu i en Joan sou part del mateix equip. Surt una fitxa que demana la capital d'Itàlia. En Joan en aquell moment dubta si la capital d'Itàlia és Roma i t'ho diu dubtant. Tu li dius que és cert, que és Roma, la capital d'Itàlia.

You and Joan are playing Trivia and you are both part of the same team. The card you get asks what is the capital of Italy. At that moment, Joan is unsure whether the capital of Italy is Rome and, doubting, he asks you whether it is or not. You tell him that it is true, that it is Rome, the capital of Italy.

2.3. Procedure

Participants were presented with one card at a time, in a random order, and were asked, for each card, to read it carefully, to imagine themselves in the situation described in the DCT and, finally, to provide an appropriate verbal response. All of the DCTs used in this experiment were designed to elicit a declarative sentence expressing confirmation.\(^9\) It is important to emphasize that participants were not asked to gesture at any point, nor to produce utterances in an "expressive" manner. Therefore, all head nod gestures obtained in this study were completely spontaneous, even though the setting of the experiment was not that of a natural conversation. Crucially, responses had to include the target word (the

---

\(^8\) The cards in the experiment included only the Catalan version. See Appendix for the full list of materials.

\(^9\) Building on the tight correlation between confirmation contexts and head nods, as described in Poggi et al. (2011: 2-4). This method has proven efficient in eliciting head nods in a pilot experiment.
capital city’s name) in a declarative sentence expressing confirmation. In cases where participants provided a response that did not include the target word (e.g., Sí, sí, és veritat. ‘Yes, yes, it’s true’), they were asked to provide another response using the name of the capital city within the sentence.

Participants were audiovisually recorded using a Panasonic HD AVCCAM, at 50 frames per second. The camcorder was placed on a tripod at a distance of approximately 1m from the participant, and its height was adjusted to the participant’s height in such a way that the recording area included the participant’s upper body and head. The participants were recorded while standing up and were asked not to hold the DCT cards while providing a response. The entire procedure lasted approximately 15 minutes.

2.4. Coding

Out of the 275 trials (11 participants x 25 items per participant), participants produced an overall of 155 head nod gestures that we were able to analyze for the purposes of this study. The remaining 120 trials were excluded from our analysis for various reasons; Table 2 provides a complete description of the distribution of the recorded data.

<table>
<thead>
<tr>
<th>Category</th>
<th>Quantity</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>head nods co-occurring with a target word</td>
<td>155</td>
<td>56.4</td>
</tr>
<tr>
<td>unassociated repetitive head nods</td>
<td>39</td>
<td>14.2</td>
</tr>
<tr>
<td>no head movement</td>
<td>30</td>
<td>10.9</td>
</tr>
<tr>
<td>head nods co-occurring with a non-target word</td>
<td>27</td>
<td>9.8</td>
</tr>
<tr>
<td>non-vertical head movement (e.g., head shake/tilt)</td>
<td>18</td>
<td>6.5</td>
</tr>
<tr>
<td>erroneous trial (target word was mispronounced)</td>
<td>3</td>
<td>1.1</td>
</tr>
<tr>
<td>not performed (accidentally forgotten)</td>
<td>3</td>
<td>1.1</td>
</tr>
<tr>
<td>total</td>
<td>275</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2: The distribution of all recorded trials.

The 155 trials that contained a head nod co-occurring with the target word were annotated in terms of speech and gesture. For speech annotation, we used the Praat software package (Boersma & Weenink 2012). We annotated start and end times for the entire utterance (Tier 1 in Figure 1), the target word (Tier 2 in Figure 1), the nuclear syllable and the pre- and post-nuclear syllables (Tier 3 in Figure 1). In addition, the point of

---

10 In those cases, the head nod usually co-occurred with the copular verb és ‘is’, preceding the target word.
intensity peak\textsuperscript{11} was also annotated for every utterance, as well as pitch accents and phrase boundary tones (Tier 4 in Figure 1). Figure 2 provides an example of the acoustic labeling in Praat, where an intensity peak within the boundaries of the accented syllable is clearly visible. The Praat annotation files where then imported into the ELAN software package (Lausberg & Sloetjes 2009) to annotate the head nod gestures.

\begin{center}
\textbf{Figure 1: Example of ELAN labeling.}
\end{center}

\textsuperscript{11} In addition to the intensity peaks, fundamental frequency peaks were also annotated when they appeared (i.e., when the target word was pronounced with a H\textsuperscript{*} or a L+H\textsuperscript{*} pitch accent). See Section 3.3 for a comparison of the correlation between the head nod apex and these two measurements (intensity and fundamental frequency peaks).
Figure 2: Example of the acoustic labeling in Praat of the target word 'Paris'.

Since, to the best of our knowledge, there are no clear conventions on how to annotate the different phases of head nods — unlike with manual gestures, where McNeill’s (1992) guidelines are conventionally used in the literature — we decided to annotate the head nod gestures in the following manner, trying to keep as much as possible to the conventions used with manual gestures. Following the definitions of McNeill (1992: 78-82), Cassell & Prevost (1998: 16-17) and Poggi et al. (2011: 2), we classified the confirmatory head nod gesture as metaphoric and triphasic. It is metaphoric since it conveys the abstract meaning of confirmation/agreement and, according to Cassell & Prevost (1998: 16), metaphoric gestures "are produced in three temporal phases". The division of the confirmatory head nod into three phases is also supported by Poggi et al. (2012: 2), who define the head nod as "a vertical head movement in which the head, after a slight tilt up, bends downward and then goes back to its starting point". Therefore, we annotated a preparation phase, a stroke and a retraction phase for each of the 155 confirmatory head nod gestures analyzed in this study. These annotations were made using two tiers in ELAN, in which four points were annotated (see Figures 3+4): (1) the beginning of the preparation phase, (2) the highest point of the head before the downward movement, (3) the lowest point of the head after the downward movement, and (4) the end of the retraction phase. The preparation phase was taken to be the interval of time during which the head was moving towards point (2), its highest point just before the downward movement. The timing of the starting point of the preparation phase (i.e., point 1) was
arrived at by examining the head movement, frame by frame, from point (2) backwards (chronologically), until the trend of the head movement began to change. The point at which a single trend movement towards point (2) began was taken to be point (1). Point (3) was annotated as the first frame in which the downward movement stopped and the head was at its lowest position after point (2). Whether the head remained in that position for several frames or not, the first of these frames was taken to be point (3). In a similar manner to the selection of point (1), point (4) was taken to be the point at which the trend of the head movement after point (3) began to change. For the sake of simplicity and conventionality, we refer to the interval between points (2) and (3) as the stroke (since this is the meaningful part of the gesture) and to point (3) in particular as the apex (since this is the point at which the head reaches its kinetic goal).

The gesture stroke is, therefore, the interval of time annotated in Tier 5 in Figure 1, i.e., the time interval between points (2) and (3). The interval annotated in Tier 6 in Figure 1 denotes the interval of time between points (1) and (4). Figures 3+4 demonstrate these four points in time.

Importantly, the recorded responses analyzed in this study had either of the following structures: (a) in 141 (91%) trials, participants located the name of the capital city at the end of the phrase (e.g., Sí, sí, la capital de França és París. 'Yes, yes, the capital of France is Paris'), (b) in 6 (3.8%) trials, the name of the capital city was in a left-dislocated position (e.g., Sí, sí, és París, la capital de França. 'Yes, yes, it is Paris, the capital of France'), and (c) in 8 (5.2%) trials, the name of the capital city was placed at the end of the phrase and the name of the state/autonomous community/district was left unpronounced (e.g., Sí, sí, és París. 'Yes, yes, it's Paris'). In either case, the target word always immediately preceded a prosodic boundary. This was crucial for us to be able to analyze the effect of the prosodic boundary on the alignment of the head nod gestures, given the differences in the metrical structures of the target words.
3. RESULTS

In this study, we wished to answer the following research questions: (a) what is the effect of the position of the accented syllable with respect to an upcoming prosodic boundary on the position of the intensity peak within the accented syllable? (b) what is the effect of the position of the accented syllable with respect to an upcoming prosodic boundary on the position of the head nod apex? and (c) does the timing of the intensity peak correlate with the timing of the head nod apex? Following Esteve-Gibert & Prieto (2013, 2014) and Nobe (1996), our predictions were that the proximity of the accented syllable to a prosodic
boundary would influence both the timing of the intensity peak and that of the head nod apex and, in addition, that these two temporal measurements would correlate.

This section is therefore divided into three subsections, one for each research question.

3.1. Timing of the intensity peak

A one-way analysis of variance (ANOVA) was performed with the metrical structure of the target word as the independent variable (five levels: monosyllabic, iambic, trochaic, amphibrach and dactylic) and the distance in milliseconds between the intensity peak and the end of the accented syllable as the dependent variable. The statistical analysis revealed that the timing of the intensity peak within the accented syllable was significantly different depending on the metrical structure of the target word, $F(4, 150) = 105.659, p < .001$. A Tukey post-hoc analysis showed that the timings of intensity peaks with respect to the end of the accented syllable varied significantly between both monosyllables and iambics with respect to the other metrical structures (all $p < .001$). There were no significant variations in timings of intensity peaks with respect to the end of the accented syllable between trochees and dactyls ($p = .897$), between trochees and amphibrachs ($p = 1.000$) or between dactyls and amphibrachs ($p = .822$). Finally, a smaller, though marginally significant, difference was found between monosyllables and iambics ($p = .045$).

Figure 5 shows the time (in milliseconds) between the intensity peaks and the end of the accented syllable for the five different metrical structures analyzed. These results are related to the findings by Esteve-Gibert & Prieto (2013) in that intensity peaks are retracted in both monosyllables and iambics (i.e., metrical structures in which the accented syllable is adjacent to the phrase boundary), while they are more closely aligned to the end of the accented syllable in trochees, dactyls and amphibrachs (i.e., metrical structures in which at least one syllable separates between the accented syllable and the phrase boundary). These results, however, do show a slightly different pattern than that presented in Esteve-Gibert & Prieto (2013). While their results showed no significant difference in the timings of intonational peaks with respect to the end of the accented syllable between monosyllables and iambics, the results obtained in the current study show a marginally significant difference between these two metrical structures with respect to the research question. However, the two sets of data might not be comparable. While Esteve-Gibert &
Prieto analyzed the timings of intonational peaks (i.e., peaks of F0 frequency), the current study analyzes the timings of intensity peaks. Unfortunately, our data did not contain enough cases of rising pitch accents for us to perform an ANOVA with only these cases, with respect to the first research question. Nevertheless, as mentioned above, Esteve-Gibert & Prieto’s results are compatible with this study’s results to the extent that both monosyllables and iambs show a significant temporal retraction of the intensity/intonation peaks with respect to the end of the accented syllable.

**Figure 5:** Box plots of the time (in milliseconds) between intensity peaks and the end of the accented syllable, separated by the prosodic word types.

---

12 Of the 155 trials analyzed in this study only 23 were pronounced with rising pitch accents. Of those 23 cases, only 2 consisted of a monosyllabic target word, and 6 consisted of an iambic target word. Therefore, comparing the means of the pitch accent timings between monosyllabic and iambic words in our data would not lead to valuable statistical insights.
3.2. Timing of the head nod apex

A one-way ANOVA was performed with the metrical structure of the target word as the independent variable (five levels) and the distance in milliseconds between the head nod apex and the end of the accented syllable as the dependent variable. The statistical analysis revealed that the timing of the head nod apex with respect to the end of the accented syllable varied significantly, depending on the metrical structure of the target word, \( F(4, 150) = 76.354, p < .001 \). A Tukey post-hoc analysis showed that timings of head nod apexes with respect to the end of the accented syllable varied significantly between both monosyllables and iambs with respect to the other metrical structures (all \( p < .001 \)). No significant differences were found between trochees and dactylys \( (p = .990) \), between trochees and amphibrachs \( (p = .448) \) or between dactyls and amphibrachs \( (p = .202) \). Unlike with the findings regarding the timings of the intensity peaks with respect to the end of the accented syllable from the previous subsection, no significant variation in the timings of head nod apexes with respect to the end of the accented syllable was found between monosyllables and iambs \( (p = 1.000) \).

Figure 6 shows the time (in milliseconds) between the head nod apex and the end of the accented syllable for the five metrical structures analyzed. Interestingly, Esteve-Gibert & Prieto (2013) did find a significant difference in gesture apex timing with respect to the end of the accented syllable between monosyllables and iambs. However, the differences between our findings and Esteve-Gibert & Prieto’s findings might stem from differences in the experimental design. While the utterances in Esteve-Gibert & Prieto (2013) always started with the target word (i.e., contained no phonological material before the target word), the utterances in our experiment always contained some phonological material prior to the target word. The existence of phonological material preceding the target word may account for the fact that no significant differences were found in head nod apex timings with respect to the end of the accented syllable between monosyllables and iambs.
Figure 6: Box plots of the time (in milliseconds) between head nod apexes and the end of the accented syllable, separated by the prosodic word types.

3.3. Intensity-gesture synchrony

The last research question in this study deals with the identity of the phonological element that co-occurs in time with the head nod apex. Following Esteve-Gibert & Prieto (2013), it would have been the intonational peak within the nuclear syllable. However, since our experimental design did not aim at obtaining contrastive focus utterances, most of our data did not contain intonational peaks within the nuclear syllable. Therefore, we decided, following Nobe (1996), to test the correlation between the head nod apexes against a related measure, namely, the intensity peak. A Pearson correlation analysis was performed.

13 In a pilot experiment conducted for this study, we used DCTs containing contrastive focus contexts, in order to obtain rising pitch accents on the target words. However, this method has proven less efficient in obtaining spontaneous head nods than the method eventually used (i.e., with confirmatory rather than contrastive contexts).
with two independent variables: the timing of the head nod apex with respect to the end of the accented syllable and the timing of the intensity peak with respect to the end of the accented syllable. This analysis revealed that the distance in time between the head nod apexes and the intensity peaks with respect to the end of the accented syllable were significantly correlated, $r(155) = .831$, $p < .001$. As in Esteve-Gibert & Prieto (2013),\(^{14}\) two additional correlation analyses were carried out to investigate the possibility that other parts of the gesture and/or the speech correlate better, as suggested in previous literature. The first additional correlation analysis was performed with the stroke onset (rather than the apex) and the intonational peak, both with respect to the end of the accented syllable as independent variables. The analysis revealed a significant correlation, though with a slightly lower correlation coefficient, $r(155) = .818$, $p < .01$. The second additional correlation analysis was performed with the head nod apex and the end of the accented syllable (rather than the intensity peak), both with respect to the onset of the accented syllable as independent variables. The analysis revealed no significant correlation, $r(155) = .096$, $p = .235$. Finally, we conducted another correlation analysis, using the 23 cases in our data that contained intonational peaks, to see if their timings correlated with the timings of the intensity peaks. A Pearson correlation analysis with the timing of the intonational peak with respect to the end of the accented syllable and the timing of the intensity peak with respect to the end of the accented syllable revealed that these two measurements were highly correlated, $r(23) = .920$, $p < .001$. These results support Esteve-Gibert & Prieto (2013)’s results, namely, that the gesture apex timing correlates with the intensity/intonational peak better than with any other phonological event.

Figures 7+8 illustrate the correlation between the head nod apex and the intensity peak as well as their temporal retraction in words with a final stressed syllable with respect to words with a penultimate stressed syllable. The head movement is represented as produced in its vertical axis. The retracted position of both multimodal events is visible when comparing iambs (illustrated in the upper graph) with trochees (the lower graph).

\(^{14}\) Esteve-Gibert & Prieto (2013) performed three additional correlation analyses, following previous literature. However, since the definition of the gesture stroke in this study is different than the definition in Esteve-Gibert & Prieto’s study, the correlation analysis that compares the end of the stroke to the intonational peak was not performed in this study, since according to our definitions for the head nod stroke, its end is the apex itself.
4. DISCUSSION

This study aimed at answering three research questions: (a) what is the effect of the position of the accented syllable with respect to an upcoming prosodic boundary on the position of the intensity peak within the accented syllable? (b) what is the effect of the position of the accented syllable with respect to an upcoming prosodic boundary on the position of the head nod apex? and (c) does the timing of the intensity peak correlate with that of the head nod apex? To answer those questions, we analyzed 155 instances of head nods co-occurring with target words in different metrical structures produced by adult Catalan speakers. The experimental design mainly followed that of Esteve-Gibert & Prieto (2013), but with several differences. First, we preferred a less controlled setting, in order to
see whether spontaneously occurring head nod gestures behave in a similar manner to the manual deictic gestures produced in the controlled setting in Esteve-Gibert & Prieto (2013). Therefore, we chose to use DCTs in order to obtain semi-spontaneous responses from our participants. Second, to obtain spontaneous head nods we opted for using confirmatory contexts in the DCTs, rather than contrastive contexts.\textsuperscript{15} This proved worthy as all instances of head nod gestures analyzed in this study were spontaneously produced. However, focusing on confirmatory statements meant that we could not rely on having a fundamental frequency peak within the nuclear accented syllable. Therefore, we decided, along the lines of Nobe (1996), to use the intensity peak as a temporal reference point, instead of the intonational peak. This decision also proved fruitful since results confirmed that the intensity peaks behaved similarly to the intonational peaks, as described in Esteve-Gibert & Prieto (2013).

Statistical analyses of our data revealed that head nod apexes and intensity peaks behaved similarly with respect to their position within the accented syllable, that is, both of them were retracted when immediately preceding a prosodic boundary. In addition, the analyses showed that the head nod apex timings correlated better with the intensity peak timings than with any other measurement within the accented syllable. We now discuss these findings.

The first main finding was that both head nod apexes and intensity peaks behave similarly regarding their temporal alignment within the accented syllable. When the accented syllable was not immediately followed by a prosodic boundary (i.e., in trochaic, dactylic and amphibrach target words), both head nod apexes and intensity peaks were closely aligned to the end of the accented syllable. However, when the accented syllable was immediately followed by a prosodic boundary (i.e., in monosyllabic and iambic target words), both head nod apexes and intensity peaks were temporally retracted.

As Esteve-Gibert & Prieto (2013) explain, the temporal retraction of intonation peaks within the accented syllable in phrase-final positions before L\% boundary tones is a well documented phenomenon in Catalan and other languages. The accepted explanation in the

\textsuperscript{15} See fn. 13.
literature for this phenomenon is that the rising pitch movement is temporally constrained by two phonological events: (a) it has to reach its peak within the nuclear accented syllable, and (b) it has to descend back to the low tone before the prosodic boundary is reached. Similarly, the same phenomenon seems to be occurring with gestures. When they are produced in a non-phrase-final position, the gesture apex is closely aligned to the end of the accented syllable (as does the intonational peak), but when the accented syllable with which they co-occur is in a phrase-final position, they are temporally retracted, so that the retraction phase can be accommodated within the phrase boundaries. Our results corroborate the findings of Esteve-Gibert et al. (2013, 2014) in this respect and, in fact, expand on them, by analyzing the alignment of head nod gestures of two additional metrical structures (i.e., dactyls and amphibrachs).

This finding therefore answers research questions (a) and (b). Both intensity peaks and head nod apexes are affected by upcoming prosodic boundaries in that both measurements are temporally retracted from the end of the accented syllable when immediately preceding a prosodic boundary.

The second main finding was that the timings of the intensity peak and of the head nod apex were the two measurements in speech and gesture with the highest correlation. This finding confirms the hypothesis that the intensity peak and the head nod apex are temporally aligned. Previous studies have suggested different alignment strategies for gesture and speech, relying on the principle that prominence in gesture correlates with prominence in speech. According to our results—which are related to those of Esteve-Gibert & Prieto (2013)—, head nod apexes seem to correlate with intensity peaks better than with any other phonological measurement within the nuclear accented syllable.

By answering these three research questions, we were able to reinforce Esteve-Gibert & Prieto’s (2013) proposal that prosodic structure influences gesture timing in two principled ways, namely, that both gesture apexes and intensity/intonational peaks co-occur with the accented syllable regardless of its position with respect to the phrase boundary, and also that these measurements are temporally retracted (within the boundaries of the accented syllable) in the presence of an upcoming prosodic boundary.
The main contribution of this study is in adding empirical data to the field of gesture-speech coordination from the point of view of head gestures which, to this day, have been studied by far less than manual gestures. Importantly, we found clear parallelisms between head and manual gestures with respect to their alignment with speech and to the influence of prosodic structure on them. In addition, we have shown that intensity peaks behave like intonational peaks, in that they are similarly affected by the prosodic structure.

For a complete understanding of the gesture-speech coordination, future research is needed. As mentioned above, the field of head gestures with respect to their temporal alignment with speech is yet to be thoroughly studied. In addition, production and perception experiments investigating the development of gesture-speech alignment during language acquisition are necessary in order to have a complete view of the importance of the alignment of these two modalities for language comprehension in particular, and for human communication in general.\[16\]

**References**


\[16\] See Esteve-Gibert & Prieto (2014) and Esteve-Gibert, Prieto & Pons (2014) for recent studies on the development of gesture-speech coordination in infants.


APPENDIX

The full list of DCT materials used in our experiment (in Catalan); the target words analyzed for the gesture-speech synchrony are given in bold face:

1. Esteu jugant al Trivial i tu i en Joan sou part del mateix equip. Surt una fitxa que demana la capital d’Itàlia. En Joan en aquell moment dubta si la capital d’Itàlia és Roma i t’ho diu dubtant. Tu li dius que és cert, que és **Roma**, la capital d’Itàlia.

2. Esteu jugant al Trivial i tu i la Montse sou part del mateix equip. Surt una fitxa que demana la capital de Senegal. La Montse en aquell moment dubta si és Dakar i t’ho diu dubtant. Tu li dius que és cert, que és **Dakar**, la capital de Senegal.

3. La Maria i tu esteu parlant de la possibilitat d’anar de viatge a Letònia aquest estiu. La Maria vol buscar vols a Letònia però no està segura si la capital de Letònia és Riga. Tu li dius que és cert, que és **Riga**, la capital de Letònia.

4. El Dani i tu esteu parlant de la possibilitat d’anar de viatge a Mònaco aquest estiu. El Dani vol buscar vols a Mònaco però no està segur si la capital de Mònaco també es diu Mònaco. Tu li dius que és cert, que és **Mònaco**, la capital de Mònaco.

5. La Marta té un amic veneçolà i vol anar a visitar-lo aquest estiu. Ella recorda que el seu amic viu a prop de la capital, però ara mateix no està segura si la capital de Veneçuela és Caracas o no. Tu li dius que és cert, que és **Caracas**, la capital de Veneçuela.

6. L’Anna té una amiga que viu a Bielorússia i vol anar a visitar-la aquest estiu. Ella recorda que la seva amiga viu a prop de la capital, però ara mateix no està segura si la capital de Bielorússia és Minsk o no. Tu li dius que és cert, que és **Minsk**, la capital de Bielorússia.

7. Tens un amic francès que està visitant Barcelona amb la seva parella. Aquest cap de setmana volen anar a Andalusia i no recorden si la capital d’Andalusia és Sevilla. T’ho pregunten dubtant. Tu els dius que és cert, que és **Sevilla**, la capital d’Andalusia.

8. Tens una amiga alemanya que està visitant Barcelona amb la seva parella. Aquest cap de setmana volen anar a Extremadura i no recorden si la capital d’Extremadura és Mèrida. Tu els dius que és cert, que és **Mèrida**, la capital d’Extremadura.

9. El Jordi i tu sou part del mateix equip que està participant en un concurs a la televisió. Us toca una pregunta que demana la capital de França. El Jordi creu que és París, però té por a equivocar-se. Tu li dius que és cert, que és **París**, la capital de França.

10. La Roser i tu sou part del mateix equip que està participant en un concurs a la televisió. Us toca una pregunta que demana la capital d’Osona. La Roser creu que és Vic, però no n’està segura. Tu li dius que és cert, que és **Vic**, la capital d’Osona.
11 Esteu jugant al Trivial i tu i l'Eva sou part del mateix equip. Surt una fitxa que demana la capital de Pallars Jussà. L'Eva en aquell moment dubta si la capital de Pallars Jussà és Tremp i t'ho diu dubtant. Tu li dius que és cert, que és **Tremp**, la capital de Pallars Jussà.

12 Esteu jugant al Trivial i tu i el Sergi sou part del mateix equip. Surt una fitxa que demana la capital de l’Alt Empordà. El Sergi en aquell moment dubta si és Figueres i t’ho diu dubtant. Tu li dius que és cert, que és **Figueres**, la capital de l’Alt Empordà.

13 La Sandra i tu esteu parlant de la possibilitat d’anar de viatge a la Xina aquest estiu. La Sandra vol buscar vols a la Xina però no està segura si la capital de la Xina és Pequín. Tu li dius que és cert, que és **Pequín**, la capital de la Xina.

14 El Xavi i tu esteu parlant de la possibilitat d’anar un cap de setmana a l’Alt Camp. El Xavi vol comprar bitllets de tren i, abans de fer-ho, et pregunta si la capital de l’Alt Camp és Valls. Tu li dius que és cert, que és **Valls**, la capital de l’Alt Camp.

15 L’Alba té una amiga peruana i vol anar a visitar-la aquest estiu. Ella recorda que la seva amiga viu a prop de la capital, però ara mateix no està segura si la capital de Perú és Lima o no. Tu li dius que és cert, que és **Lima**, la capital de Perú.

16 El Víctor té un amic que viu a Nova Zelanda i vol anar a visitar-lo aquest estiu. Ell recorda que el seu amic viu a la capital, però ara mateix no està segur si la capital de Nova Zelanda és Wellington. Tu li dius que és cert, que és **Wellington**, la capital de Nova Zelanda.

17 L’Adrià i tu sou part del mateix equip que està participant en un concurs a la televisió. Us toca una pregunta que demana la capital de Pallars Sobirà. L’Adrià creu que és Sort, però no n’està segur. Tu li dius que és cert, que és **Lima**, la capital de Pallars Sobirà.

18 El Pere té un amic que viu a Qatar i vol anar a visitar-lo aquest estiu. Ell recorda que el seu amic viu a la capital, però ara mateix no està segur si la capital de Qatar és Doha. Tu li dius que és cert, que és **Doha**, la capital de Qatar.

19 La Cristina i tu sou part del mateix equip que està participant en un concurs a la televisió. Us toca una pregunta que demana la capital de Tunísia. La Cristina creu que és Tunis, però no n’està segura. Tu li dius que és cert, que és **Tunis**, la capital de Tunísia.

20 L’Emma i tu esteu parllant de la possibilitat d’anar un cap de setmana a Castella la Manxa. L’Emma vol comprar bitllets de tren i, abans de fer-ho, et pregunta si la seva capital és Toledo. Tu li dius que és cert, que és **Toledo**, la capital de Castella la Manxa.

21 El Toni i tu sou part del mateix equip que està participant en un concurs a la televisió. Us toca una pregunta que demana la capital del Gironès. El Toni té por a equivocar-se i et pregunta si és Girona. Tu li dius que és cert, que és **Girona**, la capital del Gironès.
22 Esteu jugant al Trivial i tu i el Guillem sou part del mateix equip. Surt una fitxa que demana la capital del Ripollès. El Guillem en aquell moment dubta si la capital del Ripollès és Ripoll i t'ho pregunta. Tu li dius que és cert, que és Ripoll, la capital del Ripollès.

23 L’Àlex i tu sou part del mateix equip que està participant en un concurs a la televisió. Us toca una pregunta que demana la capital d’Afganistan. L’Àlex no està segur si és Kabul i t’ho pregunta. Tu li dius que és cert, que és Kabul, la capital d’Afganistan.

24 Esteu jugant al Trivial i tu i el Josep sou part del mateix equip. Surt una fitxa que demana la capital de l’Urgell. El Josep en aquell moment dubta si la capital de l’Urgell és Tàrrega i t’ho diu dubtant. Tu li dius que és cert, que és Tàrrega, la capital de l’Urgell.

25 El Dani i tu sou part del mateix equip que està participant en un concurs a la televisió. Surt una pregunta que demana la capital dels Estats Units. El Dani no està segur si és Washington i t’ho pregunta. Tu li dius que és cert, que és Washington, la capital dels Estats Units.