SSL

The interaction of pitch accent and gesture production in Italian and English

Diane Brentari - Giovanna Marotta
Ilaria Margherita - Angela Ott

Abstract

This article addresses the way that gesture and speech prominence are manifest in two different linguistic communities – Italian and American English. Controlled and free narratives were elicited from three adult American females (Midwestern variety) and three adult Italian females (Florentine variety) and analyzed using a method designed to measure the properties associated with pitch accents, those of gestures, and the alignment of the two together in naturally occurring speech. The results suggest some important similarities as well as some differences between American and Italian speakers. In cases where the pitch accent is not totally contained within the domain of the gesture, both groups tended to produce gestures that preceded the onset of the pitch accent, confirming previous work. Both groups overwhelmingly also used manual gestures of the hands for iconic gestures. There was a difference in the expression of prosodic gestures: while both American and Italian groups produced a greater number of prosodic gestures with the hands than the body, the Italians exhibited a stronger preference to do so. The vowels of the English speakers were also more affected by the presence of a gesture than those of the Italian speakers. The trends found in the results of this study suggest that there may be cross-linguistic differences in the properties and alignment of gestures, but larger groups of participants will be needed to conclude this with certainty, since there was a high degree of between-subject variation. In both Italian and US English, these data provide evidence for the existence of one, multi-modal system involving speech and co-speech gestures that constitutes an integrated prosodic system of a language.

Keywords: Prominence, co-speech gesture, pitch accents, audio-visual prosody.

1. *Introduction*

This research addresses the way that gesture and pitch accent (henceforth PA) are expressed in two different linguistic communities – Italian and American English. To the extent that differences are found between the two linguistic communities, this would be evidence that trends in gesturing and speech prominence are language-specific. Moreover, correlations between different aspects of gesture and speech prominence in different languages might indicate the existence of a one prosodic system involving

Received: October 2012
Accepted: February 2013

speech and the gestures that company it (co-speech gesture), and this whole multi-model system would be the one that would actually comprise an important part of the phonology of the language. Such a finding would mean that spoken language phonology research must include gestures and speech to describe the phonology of a language.

The first question of the current study concerns the timing between PA as expressed in gesture and speech. Do Italians and Americans differ in their timing of gestures with respect to their PA? The audio and gestural data will be analyzed to determine if and how gestures overlap with the PA prominences in the two linguistic communities.

The second focus of this study is the relationship between the primary articulator and the gesture type. Are iconic gestures produced by the body more often, less often, or equal to the hands? Are prosodic gestures produced more often by the hands or by the face and body? Is there a significant difference between which part of the body Americans and Italians use to produce iconic or prosodic gestures? A third area we wish to explore is the effect that a gesture might have on acoustic properties of their associated vowel, specifically vowel duration and frequency. These specific questions are all relevant to the broad question of finding out if speech prominence and gesture are interconnected in such a way that we can learn something about the larger structure of prosodic phonology by studying these relationships.

2. *Accent and the phonological phrase*

PA is defined here as a local degree of stress, prominence or emphasis (Lieberman and Pierrehumbert, 1984; Beckman and Pierrehumbert, 1986; Beckman and Venditti, 2011), which can also have an informational role of drawing attention to new information in the utterance (Ito and Sheer, 2008; Ito *et al*., 2012). Prominence falls on the nuclear syllable in the phonological phrase in both syllable-timed and stress-timed languages (Beckman and Venditti, 2011).

The phonological phrase is defined as the prosodic domain, which contains a single PA (Pierrehumbert, 1980; Ladd, 1996). This unit is part of the phonological hierarchy, which was defined and described in generative terms by Nespor and Vogel (1986), shown in (1).

(1) *Prosodic Hierarchy* (Nespor and Vogel, 1986)

Utterance → I(ntonational) Phrase → P(honological) Phrase → P(honological) Word[1]

Each unit in the prosodic hierarchy is individuated for a given language on the basis of phonological constraints that apply uniquely to that domain, making them an essential part of the phonological grammar of a given language. The following is an example of how an utterance would be divided into phonological phrases within a single Intonational Phrase, with the pitch accented syllables shown in bold (2).

(2) *Phonological phrases within a single intonational phrase, with PA indicated in bold* (Nespor and Vogel, 1986/2008: xxii)

[The oldest **boys**] [are in**vi**ted] [to **come**] [to the **pool**] [every **Mon**day]

Gussenhoven (2004) was concerned with universals and cross-linguistic studies of prosodic phonology. He outlined three ways that prominences are expressed in spoken language. Following Ohala's work (1984) on intonation, Gussenhoven (2004) describes three phonological "codes" that can be used to express PA by setting the accented syllable apart from the non-accented ones. The *Frequency Code* (pitch) concerns the use of pitch, the *Effort Code* (duration) relates to muscular effort – e.g., stronger syllable contacts. The *Production Code* (intensity), is concerned with air pressure at the beginning of an utterance than at the end, resulting in a gradual drop in intensity across the duration of an utterance. It is thus expected that any cross-linguistic differences in the expression of PA would result from differential use of these codes. Duration, intensity and frequency may be indicators of prominence, but duration and frequency are used in PA, which is the focus of the current study.

[1] This prosodic hierarchy shown here includes only units above the level of the prosodic word. The hierarchy can be further analyzed into Foot → Syllable → Mora in models of prosodic analysis. Also, in the original hierarchy of Nespor and Vogel (1986), there is an another level just above the P-Word – Clitic Group – not shown here.

3. *Gestures*

A review of gesture literature reveals that researchers employ a variety of definitions for what gesture is and how to classify different types of gesture (McNeill, 1992, 2000, 2005; Kendon, 2004). Manual expression can take many forms – co-speech gesture, pantomime, emblems and sign languages (McNeill, 2005). Co-speech gesture, according to McNeill, does not have linguistic properties, is not conventionalized, and is arbitrary and contains meaning only as a whole entity (Loehr, 2004; McNeill, 2005).

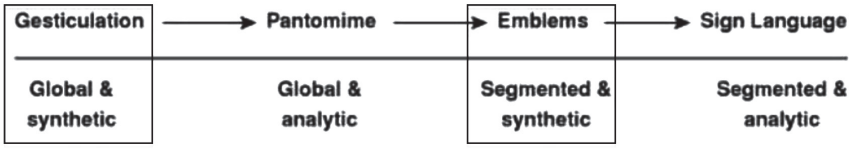

Figure 1. *Semiotic categories of manual expression (McNeill, 2005: 10). Those included in this study are "gesticulation" and "emblems".*

The definition of "gesticulation" in Figure 1 includes only co-speech gesture and excludes any gesture used without speech; these are *representational gestures* and *beats,* and are produced by the hands, eyebrows, face, and body. Representational gestures include iconic, metaphoric, and emblematic gestures that contribute meaning to a message. In this study representational gestures were not further subdivided. *Beat* gestures have no meaning and have often been considered most relevant for prosodic analyses (Duncan, 1999; Gullberg, 2006; McNeill, 1992, 2000); however, more recent work on the specific timing of gestures and speech suggests that all gestures have prosodically important timing (McClave, 1998; Krahmer and Swerts, 2005, 2007; Loehr, 2004, 2007). Loehr (2007) proposes that the difference between representational and beat gestures is that representational gestures have meaningful content, while beats are empty of meaningful content.

In the present study, we deal with representational co-speech gestures, grooming gestures, and emblems (though there were very few). Emblems were included because it was thought that they might also serve a prosodic purpose if they typically occurred during the pauses between phrases. Grooming gestures were also included for the same reason – i.e., they are

often produced at the end of a prosodic unit. Examples of an iconic gesture, a prosodic beat gesture, and a prosodic grooming gesture are given in Figure 2.

Figure 2. *Examples of the common gesture types used in this analysis: an iconic gesture representing Tweety holding binoculars (left); a prosodic gesture with eyebrows raised (center) and a prosodic grooming gesture – scratching one's eyebrow (right).*

4. *The interaction of speech prominence and gesture*

Speech prominence and gesture are interrelated in many ways. As for prominence in speech, recent empirical research on speech perception has focussed on the perceptual effects produced by the interaction of the two leading acoustic parameters, i.e. duration and F0, although we know that the perceptual salience of an auditory stimulus depends on the peculiar combination of different physical elements (e.g. frequency, duration, intensity, voice quality), and is not simply derivable by a single one of them (Niebuhr, 2009).

Prominent vowels can be defined as segments having a special degree of perceptive salience in an utterance. As is well known, in a phonetic string, a segment as well as a syllable, can be perceived as prominent after a relevant modification of the three basic acoustic parameters, i.e. duration, intensity and frequency, which are reflected on the perceptual side by changes in length, volume and tone (cf. Rietveld and Gussenhoven, 1985; Kohler, 2008). However, the relative role played by these physical parameters in the perception of prominence is still unclear; therefore, studying the special relationship between duration and F0 is not an easy task. For instance, listeners normally perceive modulated vowels as longer; at the same time, a longer vowel is normally perceived as prominent (Gussenhoven, 2004). Moreover,

the effects of gesture on the perception of prominence is a relatively new area of research (Krahmer and Swerts, 2005, 2007; Foxton, Riviere and Barone, 2010; Cvejic, Kim and Davis, 2012).

One way to approach the perception of prominence by human beings could be to study the relationship between prominence in speech and production of gestures; in other words, to study whether gesture has an effect on how humans perceive prominence.

Munhall *et al*. (2004) studied the effect of head movements on speech perception. Twelve native Japanese speakers watched a video of a talking-head accompanied by noisy speech. The participants correctly identified more speech syllables when the talking-head moved naturally (without synthetic manipulation) in concert with the speech. This study shows how closely connected gesture and speech are in listener perception, and how the presence of a beat gesture enhances perception.

Krahmer and Swerts (2007) also looked at the relationship between acoustic cues and visual cues by noting acoustic effects of gesture in production and perception of speech. They asked, «Does producing a visual beat affect how prominence is realized in speech?» and «Do beats change how prominences are perceived?» (Krahmer and Swerts, 2007: 396). Their definition of prominence was «the relative accentual strength with which words are realized in a spoken utterance» (Krahmer and Swerts, 2007: 396). They also investigated whether specific articulators (eyebrows, head, and hands) impact the speech properties in production. They investigated the effect of beat gestures on prominent syllables in Dutch speakers. Participants were instructed to produce a target sentence containing two possible target words with the vowel /a/. The authors found that prominent vowels accompanied by a beat gesture were produced with more prominent acoustic features than those without a gesture. Duration, F0, F2, and intensity were affected. Prominent syllables accompanied by a gesture were also more likely to be perceived as prominent. Krahmer and Swerts' (2007) data showed a clear relationship between PA and gestures for the speaker. This study will build on Krahmer and Swerts' work using a more natural setting, including a personal, first person narrative and a narrative told in the third person. Krahmer and Swerts suggest that further studies with spontaneous speech would be useful in understanding more about the relationship between prominence and gesture. They also suggest that results may be culture or language dependent (Krahmer and Swerts, 2007: 411). This study will address some of these questions.

A third important issue could be in the timing of the phases of gesture to PA. Concerning the timing of speech and gesture, one early view was that gestures were more likely to appear with pauses than with the speech (Beattie and Aboudan, 1994). This early work used data from dialogues where speakers were interrupted. Nobe (1996, 2000) disproved these results by pointing out that interrupted speech may not include typical timing of gesture with speech, and in more typical contexts he found that gestures were more likely to co-occur together with speech. He also introduced a gesture and acoustic-peak synchrony rule, which states "The gesture onset and stroke (see below) precede and/or co-exist with, but do not start after, the (co-occurring) peak of F0 and intensity." In a subsequent study by McClave (1998) it was found that the gestural stroke co-occurred with the final stressed syllable and with the most significant change in pitch in just over half of the intonation groups examined. The stroke co-occurred with another stressed syllable in another 25 percent of the groups. McClave found that the data from her study supported Nobe's rule; results from the present study may also be able to confirm this claim. The present study is intended to build on the work of McClave (1998) and Krahmer and Swerts (2007) by considering speakers from different linguistic communities.

The timing of gesture and speech was further investigated using the three *phases* of a manual gesture. The "preparatory phase" is a movement of the hand or other articulator from rest position to the point where the "stroke" begins. The stroke is main part of the gesture articulatorily (McNeill, 1992: 13). The "retraction phrase" is the movement towards "resetting" the speaker's neutral position.

One way to think about these phases is to imagine someone gesturing that they are throwing a ball. A hand would be in a closed-fist shape, the elbow would bend, and the hand would come up to the shoulder; this action would constitute the preparation phase. Throwing the imaginary ball would be the stroke, the main part of the gesture. Retraction would be the hand coming out of the hand shape and resting back on the speaker's lap. Loehr (2004) investigated these claims in phonology by tracking the «apex of movement», a single instant in time at the peak of the stroke, or the «kinetic goal» of the stroke (2004: 89). Loehr found that «apexes of gestural strokes and pitch accents aligned consistently, and gestural phrases and intermediate phrases aligned quite often» (2004: iii). To count as "near" the PA, Loehr used a window of 275 milliseconds before and after the stroke. The current study also uses a window around the stroke as defining the near-

ness of the gesture to the PA, but in the current study the onset and offset of the stroke and of the whole gesture were considered, rather than employing the apex as the sole moment of comparison.

5. *Methods*

5.1. *Participants*

Participants were recruited by advertisement at the University of Florence and Purdue University. Three native American English speakers and three native Italian speakers participated in this study. All participants were female and between 18-30 years of age. Each participant was paid for their participation in the study (IRB Protocol 0603003706 to Brentari and colleagues). Given the small number of participants we included only young females to rule out that any differences that might be found were gender-based or age-based.

5.2. *Stimuli and procedures*

Two recordings were collected from each participant (three American English speakers and three Italian speakers): the third person narrative was a retelling of a Sylvester and Tweety cartoon (McNeill, 1992) and the first person narrative was about an event that happened in their life. The participants were seated in front of a camera with a laptop at their side. The cartoon was first shown in its entirety and then a second time episode by episode after which the subject re-told the episode. Participants were instructed to describe what they saw in the cartoon. The video and audio recordings of the cartoon were divided electronically into seven episodes. For the life event, subjects were instructed to tell a short narrative of up to 5 minutes about an event in their life. The interlocutors were native speakers of the target language, except in one Italian case where the interlocutor was one of the authors of the study. This resulted in a speech sample of 36'05". Across subjects the Tweety and Sylvester narratives averaged 8'16" and the life event narratives 4'25" in length.

5.3. *Coding procedures*

The audio files containing the PA were analyzed using PRAAT (Boersma and Weenink, 2009) and the video files containing the participants'

gestures were annotated using ELAN (EUDICO Linguistic Annotator), a tool developed at the Max Planck Institute for Psycholinguistics, Nijmegen, for the analysis of speech, sign language, and gesture (Lausberg and Sloetjes, 2009). The set of nine tiers used in this analysis are described in detail below. Reliability among transcribers (the primary transcriber and two additional transcribers) for vowel duration and for the ELAN coding, described below, was 90%. The transcription included tiers for the transcribed text (prominent vowels were transcribed in IPA), tiers that described gesture properties, tiers that described properties of the PA[2], well as tiers that described the relation between the two.

5.4. *Pitch accent and prominence*

A transcript of the participants' responses was created. Three native American English and three native Italian judges identified prominences in each of the participants' files in their respective language. All judges had taken at least one course in linguistics and understood the notion of prominence and PA from the phonological and informational perspectives. Only the prominent vowels that had agreement between at least two judges were further analyzed in PRAAT[3]. A script (based on Crosswhite, 2009) was created to take fundamental frequency, duration, and formant measurements of the annotations in the phrase and prominence tiers. Duration of the vowel was measured from the clear appearance of F0, F1, F2, and F3 until they began to degrade. The frequency values for F0, F1,F2, and F3 were selected at the midpoint of the production of the vowel production. Once the script was run, the prominent vowels accompanied by a gesture were analyzed separately from those that were produced without a gesture then divided into five vowels groups in English and Italian /i, e, a, o, u/.

These five vowels were analyzed because only these five had at least one token in both languages that was accompanied by a gesture as well as one that was not accompanied by a gesture. This resulted in 923 pitch-accented vowels that were included in the analysis. The other vowels in English were not analyzed since they had no Italian equivalents[4]. The properties of the PA

[2] In particular, F0 contour, scaling and ToBI annotation.

[3] It is worth noting that a couple of the Italian judges identified two levels of prominence – one type more prominent, and another, more subtle.

[4] We are perfectly conscious that in such way we run the risk to ignore some prominent vowels of the English speech.

that met these criteria were imported from PRAAT text grids for further analysis.

5.5. *Gestures*

Gestures were also according to a number of descriptive tiers, which were ultimately grouped together into increasingly more general tiers for analysis. Our system is shown in Figure 3, which indicates the more descriptive tiers at the bottom and the more general ones at the top.

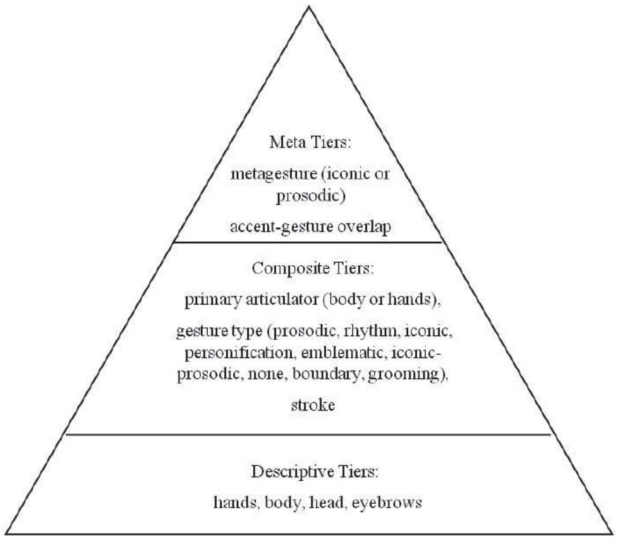

Figure 3. *Coding of gesture and overlap with the PA in ELAN.*

The bottom section of tiers included *descriptive tiers*. Annotations the content of each gesture with respect to each articulator were included for: *hand, head*, *eyebrows* and *body*. A notation system for handshape developed by Eccarius and Brentari (2008) was utilized for the coding of handshape. Eyebrows were noted as either brows up or brows down. Head tilts and body tilts were transcribed when that part of the body moved forward, backward, or to the side from a neutral position identified for each participant and each narrative (cf. Brentari, Nadolske and Wolford, 2012).

The second level in Figure 3 shows three *composite tiers*, which categorizes the descriptive annotations from the lower level: *primary articulator,*

gesture type, and *gestural phase*. The primary articulator tier used the descriptive tiers as a guide. Any gesture on the hands was coded as "hands", and any gesture that was made up exclusively by the head, eyebrows, or body was coded as "body"; that is, if both hands and body were used, the hands were given priority. Gesture type refers whether the gesture was representational (iconic or emblematic) or purely prosodic (grooming or beat). The three phases of manual gestures were also coded – the stroke, preparatory phase, and retraction phases.

The top level of the pyramid in Figure 3 contains the *meta-tiers*, which are tiers that categorize the information on the second level of the pyramid. The *metagesture* tier included annotations for whether the gesture was "prosodic" or "iconic" overall. And the *accent-gesture overlap* tier contained annotations for the degree and type of overlap between the PA and the gesture.

The accent-gesture overlap tier requires further explanation. The edges of the PA were considered the reference points for comparison – (I)nitial and (F)inal. The gesture might overlap with either or both of these edges. The edges of the PA might also overlap with different phases of the gesture – with the stroke, within a 150 millisecond window of one of the edges of the stroke, or with the edges of the whole gesture, including the preparation and retraction phases. The overlap might also be more or less precise with respect to the PA: exact overlap meant the corresponding edges of the PA and gesture were exactly the same (initial edge of the PA and the initial edge of the gestural domain), or the timing of the overlap could also be with respect to the *non*-corresponding edges of the pitch and gesture – initial edge of PA and with the final edge of the gestural domain. Cases of no-overlap were also coded as such – either a gesture with no overlap with the PA, or vice versa, a PA with no accompanying gesture.

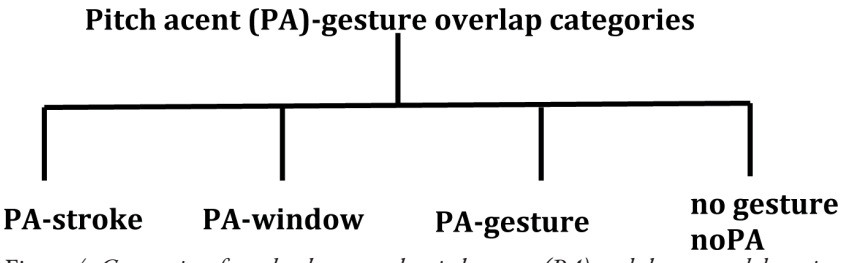

Figure 4. *Categories of overlap between the pitch accent (PA) and the gestural domain.*

6. *Results and discussion*

We report on three results from three analyses. Inferential statistical analyses were not possible with our data. There was too much inter-subject variability with only three subjects per group, so group effects did not emerge. Moreover, there were so few PAs that occurred without a gesture that a reliable comparison between the two environments was not possible. We therefore report only on the descriptive patterns of our findings. The patterns in the third person narrative (the *Tweety Bird* narrative) and the first person narrative (Life Event) were the same for each subject, so the results were combined.

There were important similarities in all of the subjects with respect to the timing of gestures and PA. Five of the six participants had a higher number of cases where the entire PA occurred within the stroke of the gesture than any other type of overlap (see Table 1). The one exception to this was Italian 1, who had a high number of cases in which the stroke ended exactly when the PA began, and these were coded as L-Overlap.

Nobe, McClave and Loehr's results on Americans were confirmed for all participants: a higher proportion of gestures overlapped with the left edge of the PA than with the right edge (see Table 1).

	L-Overlap	**R-Overlap**	**Total stroke**	**Total gesture**
American 1	0.31	0.13	0.51	0.05
American 2	0.31	0.16	0.38	0.14
American 3	0.27	0.14	0.45	0.13
Italian 1	0.40	0.17	0.26	0.16
Italian 2	0.30	0.16	0.41	0.13
Italian 3	0.31	0.12	0.40	0.17

Table 1. *Proportion of types of overlap by participant. Total overlap was the most common type of overlap (the PA totally contained within the stroke); partial overlap with the left edge of the PA was the second most common type of overlap.*

With respect to the use of the hands or body as the primary articulator, the results show the use of manual gestures (those produced by the hands) 99% and 89% of the time for iconic gestures for Italians and American participants, respectively; therefore, it is clear that gestures that contribute substantive meaning to an utterance are largely produced on the hands. The

prosodic gestures – primarily beats – showed a wide degree of inter-subject variability (see Table 2).

	prosodic gestures-hands	**prosodic gestures-body**
American 1	.15	.85
American 2	.97	.03
American 3	.78	.22
Italian 1	.91	.09
Italian 2	.76	.24
Italian 3	.98	.02

Table 2. *Proportion of prosodic gestures produced primarily by the hands vs. those produced exclusively by the body.*

Three participants – two Italians and one American – produced less than 10% of their prosodic gestures exclusively on the body – head tilts, eyebrow movements, or body tilts. The other three participants – two Americans and one Italian – produced at least 20% of their gestures exclusively on the body, and one American produced 85%. With such a small sample, it is not possible to say more, except that that there appears to be a tendency for Americans to produce more prosodic gestures exclusively on the body.

Thus far the results discussed concern timing of pitch accented vowels and gestures produced in concert. The results concerning the effect of the gesture on the acoustic properties of the vowel are also interesting. When all of the vowels are considered together no differences were evident between vowels produced with or without an accompanying gesture. However, if the five vowels are analyzed separately – [a, e, i, o, u] – we find a noticeable effect only on the vowel /a/. After the mean was calculated for each acoustic measure under investigation – duration, F0, F1, F2, and F3, proportions of the mean were calculated when the prominent vowel was accompanied or not accompanied by a gesture. Duration and F0 were the properties affected by the presence of a gesture, and /a/ vowel was more affected in both properties in the American than in Italian participants (Figure 5). A possible explanation for this effect is that /a/ is the only vowel that is not tense among the five in our study; it is typically represented by the single feature [low]. It may vary along the tense-lax continuum, resulting in more co-occurring variation in duration and F0 as well. This effect may have appeared only in English, at least in part, because English is stress-timed, while Italian is "syllable-timed".

Perhaps there is a larger principle at work about the variability of vowels in a "stress-timed" prosodic system. In other words, the same principle that allows vowels to be reduced in unstressed position in a stress-timed language can be extended to obtain a distinction related to the use of gesture. The lack of a gesture causes even pitch-accented vowels to be slightly reduced, or, conversely, that those accompanied by a gesture to be slightly fuller. If this is the case, only English vowels would be affected by the presence of a gesture in duration and F0.

As is well known, English is a stress-timed language, whereas Italian is a syllable-timed language. Therefore, in a language like English, the canonical targets are reached only in some prominent positions, whereas in unstressed position the process of articulatory and acoustic reduction are more relevant. On the other hand, in a prosodic system like Italian, the compensatory phenomena are less evident, with the effect of more stable rhythmic patterns.

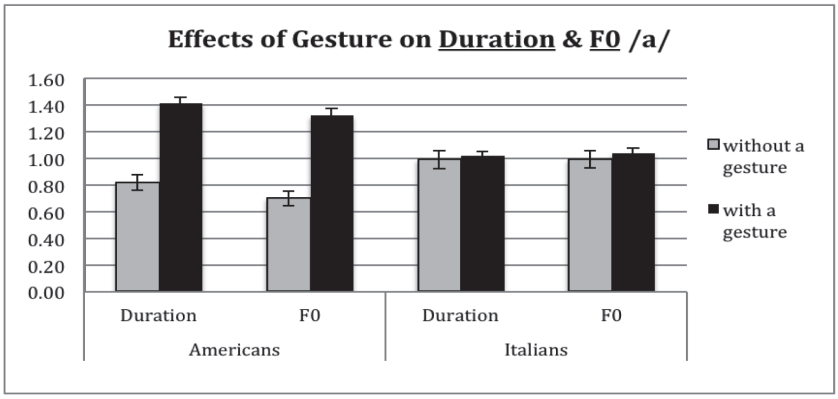

Figure 5. *The effect of the presence of a gesture on the vowel /a/ in American and Italian participants with error bars. The /a/ vowels produced by American participants were more affected by the presence of a gesture than the /a/ produced by Italian participants. The vertical axis represents the mean duration as 1.0. The bars indicate the average distance from the mean when /a/ co-occurred with a gesture and when it did not.*

7. Conclusions

This study attempted to outline a potential methodology for quantifying the relationship between PA and gesture that is both reliable and informative. We acknowledge that our sample size was quite small, and that

the effects seen might be idiosyncratic; however, this is the first study, to our knowledge, comparing Italian and English with respect to the gestural dimensions addressed – the timing of gesture and speech, and the primary articulator used for producing prosodic gestures.

First, it was expected that gestures would precede PA and that the beginning of PA would overlap with the main part of the gesture. In relation to previous work, these results confirm two current proposals. The phonological synchrony rule is confirmed (Loehr, 2004: 169: «a gesture coincides with or slightly precedes the co-occurring word, but never follows it»); there were more instances where the stroke partially preceded the PA than followed it. In addition, Loehr found that apexes of strokes aligned with PA in his data (Loehr, 2004). Considering that an apex is within a stroke, the data reported here corroborates these earlier findings; PA tends to overlap with the stroke. Not only does the PA overlap with the main categories of stroke, there are far more instances of PA with an accompanying gesture than those that do not occur with a gesture.

An effect of gesture on the PA was expected for the acoustic properties of duration, F0, and F2. Results from the individual vowel analysis partly corroborate these earlier findings from Krahmer and Swerts (2007). Recall that they compared two accented vowels in a sentence and compared those vowels in an acoustic only and in an acoustic plus gesture condition. They found higher F0, lower F2, and longer duration of the target vowels. The results of the current study can confirm an effect of gesture on a prominence's duration (longer) and F0 (higher) for the American vowel /a/; in contrast the duration and F0 of the Italian vowel /a/ was virtually unaffected by the presence of a gesture. Perhaps this effect is seen only in /a/ because it is the only vowel that is not tense among the five vowels studied and this fact allows for more variability in duration and F0. This effect might have appeared only in English participants (not in the Italian participants) because Italian is "syllable-timed", while English is "stress-timed".

It was also hypothesized that Italians would show a stronger preference for gesturing with their hands than Americans. The results show that Italians did indeed have a stronger preference to gesture with their hands than Americans, but only when producing prosodic gestures.

In future work this method could be used to address differences concerning the timing of gesture with speech due to gender, dialect or language using larger groups of participants. Using results from production studies such as these, perception studies would be a fruitful avenue for future

research as well. For example, as differences in timing become clear, such differences could be investigated by creating stimuli misaligning the gestural and acoustic signal according to expectations and determining whether speakers notice this difference either passively, using ERP, or actively, using grammaticality judgments.

In conclusion this work is in accord with McNeill's claim that gesture and speech comprise a single, poly-modal system, each contributing its own set of meanings, and given the results here, contributing distinct information about prosody as well. In particular, our data support the hypothesis of a strong connection between prosodic features and gestures in conveying prominence. In general, this work suggests that the phonology of spoken languages includes more than speech production alone, because the prosodic system includes both speech and gesture.

Bibliography

Beattie, G. and Aboudan, R. (1994), *Gestures, pauses and speech: an experimental investigation of the effects of changing social context on their precise temporal relationship*, in «Semiotica», 99, pp. 239-272.

Beckman, M. and Pierrehumbert, J. (1986), *Intonation in Japanese and English*, in «Phonology Yearbook», 3, pp. 255-309.

Beckman, M. and Venditti, J.J. (2011), *Intonation*, in Goldsmith, J., Riggle, J. and Yu, A. (2011, eds.), *Handbook of Phonological Theory*, 2nd Edition, Blackwell, London, pp. 485-532.

Bertinetto, P.M. (1989), *Reflections on the dichotomy 'stress- vs. syllable-timing'*, in «Revue de Phonétique Appliquée», 91/93, pp. 99-130.

Boersma, P. and Weenink, D. (2009), *Praat: doing phonetics by computer* (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from *http://www.praat.org/*.

Brentari, D., Nadolske, M. and Wolford, G. (2012), *Can experience with co-speech gesture influence the prosody of a sign language? Sign language prosodic cues in bimodal bilinguals*, in «Bilingualism: Language and Cognition», 15, 2, pp. 402-412.

Crosswhite, K. (2009), *Duration Logging Script. http://web.archive.org/web/20030620172734/ling.rochester.edu/people/cross/scripts.html* (15 March 2010).

Cvejic, E., Kim, J. and Davis, C. (2012), *Recognizing prosody across modalities, face areas and speakers: Examining perceivers' sensitivity to variable realizations of visual prosody*, in «Cognition», 122, pp. 442-453.

Dauer, R.M. (1983), *Stress-timing and syllable-timing reanalyzed*, in «Journal of Phonetics», 11, pp. 51-62.

Duncan, S. (1999), *Language and Communication*, in Wilson, R.A. and Keil, F.C. (1999, eds.), *The MIT Encyclopedia of the Cognitive Sciences*, MIT Press, Cambridge (MA), pp. 439-441.

Eccarius, P. and Brentari, D. (2008), *Handshape Coding Made Easier: A theoretically based notation for phonological transcription*, in «Sign Language and Linguistics», 11, pp. 69-101.

Foxton, J., Riviere, L.-D. and Barone, P. (2010), *Cross-modal facilitation in speech processing*, in «Cognition», 115, pp. 71-78.

Gullberg, M. (2006), *Some reasons for studying gesture and second language acquisition*, in «Applied Linguistics», 44, pp. 103-124.

Gussenhoven, C. (2004), *The phonology of tone and intonation*, Cambridge University Press, Cambridge.

Ito, K. and Speer, S. (2008), *Anticipatory effects of intonation: Eye movements during instructed visual search*, in «Journal of Memory and Language», 58, pp. 541-573.

Ito, K., Jincho, N., Minai, U., Yamane, N. and Mazuka, R. (2012), *Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds*, in «Journal of Memory and Language», 66, pp. 265-284.

Kendon, A. (2004), *Gesture: Visible Action as Utterance*, Cambridge University Press, Cambridge.

Kohler, K.J. (2008), *The Perception of Prominence Patterns*, in «Phonetica», 65, pp. 257-269.

Krahmer, E. and Swerts, M. (2005), *How children and adults produce and perceive uncertainty in audiovisual speech*, in «Language and Speech», 48, 1, pp. 29-54.

Krahmer, E. and Swerts, M. (2007), *The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception*, in «Journal of Memory and Language», 57, pp. 396-414.

Ladd, R. (1996), *Intonational Phonology*, Cambridge University Press, Cambridge.

Lausberg, H. and Sloetjes, H. (2009), *Coding gestural behavior with the NEUROGES-ELAN system*, in «Behavior Research Methods, Instruments, & Computers», 41, 3, pp. 841-849 (*http://tla.mpi.nl/tools/tla-tools/elan/*).

Liberman, M. and Pierrehumbert, J. (1984), *Intonational Invariance under Changes in Pitch Range and Length*, in Aronoff M. and Oehrle, R. (1984, eds.), *Language Sound Structure*, MIT Press, Cambridge (MA), pp. 157-233.

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

Loehr, D. P. (2007), *Aspects of rhythm in gesture and speech*, in «Gesture», 7, 2, pp. 179-214.

Marotta, G. (1985), *Modelli e misure ritmiche: la durata vocalica in italiano*, Zanichelli, Bologna.

Marotta, G. (2003), *What does Phonology tell us about Stress and Rhythm? Some Reflections on the Phonology of Stress*, in Solé, M.J., Recasens, D. and Romero, J. (2003, eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Vol. 1, Universitat Autònoma de Barcelona, Barcelona, pp. 333-336.

McClave, E. (1998), *Pitch and Manual Gestures*, in «Journal of Psycholinguistic Research», 27, pp. 69-90.

McNeill, D. (1992), *Hand and Mind*: *What Gestures Reveal about Thought*, University of Chicago Press, Chicago.

McNeill, D. (2000), *Language and Gesture*, Cambridge University Press, Cambridge.

McNeill, D. (2005), *Gesture and Thought*, University of Chicago Press, Chicago.

Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T. and Vatikiotis-Bateson, E. (2004), *Visual Prosody and Speech Intelligibility; Head Movement Improves Auditory Speech Perception*, in «Psychological Science», 15, pp. 133-137.

Nespor, M. and Vogel, I. (1986), *Prosodic Phonology*, Foris, Dordrecht (republished: 2008, Mouton De Gruyter, Berlin).

Niebuhr, O. (2009), *F0-based rhythm effects on the perception of local syllable prominence*, in «Phonetica», 66, pp. 95-112.

Nobe, S. (1996), *Representational gestures, cognitive rhythms, and acoustic aspects of speech: a network threshold model of gesture production*, Doctoral dissertation, University of Chicago.

Nobe, S. (2000), *Where do* most *spontaneous representational gestures actually occur with respect to speech?*, in McNeill, D. (2000, ed.), *Language and Gesture*, Cambridge University Press, New York, pp. 186-198.

Ohala, J. (1984), *An ethological perspective on common cross-language utilization of F0 of voice*, in «Phonetica», 41, pp. 1-16.

Pierrehumbert, J. (1980), *The phonology and phonetics of English intonation*, PhD dissertation, MIT.

Rietveld, A.C.M. and Gussenhoven, C. (1985), *On the relation between speech excursion size and pitch prominence*, in «Journal of Phonetics», 13, pp. 299-308.

Diane Brentari
Department of Linguistics
University of Chicago
1010 East 59th Street
Chicago, IL 60637 (USA)
dbrentari@uchicago.edu

Giovanna Marotta
Dipartimento di Filologia, Letteratura e Linguistica
Università di Pisa
via Santa Maria 36
56126 Pisa (Italy)
gmarotta@ling.unipi.it

Ilaria Margherita
Dipartimento di Filologia, Letteratura e Linguistica
Università di Pisa
via Santa Maria 36
56126 Pisa (Italy)
ilariamargherita@libero.it

Angela Ott
Linguistics Program
Purdue University
100 North University St
West Lafayette, IN 47907 (USA)
otta@purdue.edu