frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 07 July 2014
doi: 10.3389/fpsyg.2014.00700

Prosody in the hands of the speaker

Bahia Guellaï[1]*, Alan Langus[2] and Marina Nespor[2]

[1] Laboratoire Ethologie, Cognition, Développement, Département de Psychologie, Université Paris Ouest Nanterre La Défense, Nanterre, France
[2] Language Cognition and Development Laboratory, Cognitive Neuroscience Sector, International School for Advanced Studies, Trieste, Italy

Edited by:
Iris Berent, Northeastern University, USA
Susan Goldin-Meadow, University of Chicago, USA

Reviewed by:
Wendy Sandler, University of Haifa, Israel
Diane Lillo-Martin, University of Connecticut, USA

***Correspondence:**
Bahia Guellaï, Laboratoire Ethologie, Cognition, Développement, Département de Psychologie, Université Paris Ouest Nanterre La Défense, 200, Avenue de la République, Nanterre 92000, France
e-mail: bahia.guellai@gmail.com

In everyday life, speech is accompanied by gestures. In the present study, two experiments tested the possibility that spontaneous gestures accompanying speech carry prosodic information. Experiment 1 showed that gestures provide prosodic information, as adults are able to perceive the congruency between low-pass filtered—thus unintelligible—speech and the gestures of the speaker. Experiment 2 shows that in the case of ambiguous sentences (i.e., sentences with two alternative meanings depending on their prosody) mismatched prosody and gestures lead participants to choose more often the meaning signaled by gestures. Our results demonstrate that the prosody that characterizes speech is not a modality specific phenomenon: it is also perceived in the spontaneous gestures that accompany speech. We draw the conclusion that spontaneous gestures and speech form a single communication system where the suprasegmental aspects of spoken language are mapped to the motor-programs responsible for the production of both speech sounds and hand gestures.

Keywords: gestures, comprehension, speech perception, ambiguity, prosody

INTRODUCTION

Human language is a multimodal experience: it is perceived through both ears and eyes. When perceiving speech, adults automatically integrate auditory and visual information (McGurk and MacDonald, 1976), and seeing someone speaking may improve speech intelligibility (Sumby and Pollack, 1954). The visual information involved in speech is not limited to the lips, the mouth and the head, but can also involve other cues such as eyebrow movements (Bernstein et al., 1998; Graf et al., 2002; Krahmer and Swerts, 2004; Munhall et al., 2004). In fact, in face-to-face interactions people use more than their voice to communicate: the whole body is involved and may serve informative purposes (Kendon, 1994; Kelly and Barr, 1999 for a review). For example, when interacting with others, people all around the world usually also produce spontaneous gestures while talking. In fact gestures are so connected with speech that people may be found gesturing when nobody sees them (Corballis, 2002) and even congenitally blind people gesture when interacting with each other (Iverson and Goldin-Meadow, 1998). Yet, the role of gestures that accompany speech (i.e., co-speech gestures) in communication is still not well understood and little if any attention to the relation between co-speech gestures and the syntactic and prosodic structure of spoken language has been paid in previous studies. Some authors claim that these co-speech gestures are not produced to serve any communicative purposes (Rimé and Shiaratura, 1991). On the contrary, others suggest that gestures and speech are parts of the same system and are performed for the purpose of expression (Kendon, 1983; McNeill, 1992). One way to understand the implication of co-speech gestures in communication is to study their implications at the different levels of the utterance. The present study aimed to investigate the role of gestures that accompany speech at the prosodic level in speech perception.

Gestures accompanying speech are known to ease the speaker's cognitive load, and gesturing helps solving diverse individual tasks ranging from mathematics to spatial reasoning (Cook and Goldin-Meadow, 2006; Chu and Kita, 2011). Gestures are also believed to promote learning in adults as well as in children (Ping and Goldin-Meadow, 2010), to aid the conceptual planning of messages (Alibali et al., 2000), and to facilitate lexical access (Alibali et al., 2000). This suggests that gestures that accompany speech might maximize information about events by providing it cross-modally (de Ruiter et al., 2012). In fact, human infants' canonical babbling is temporally related to rhythmic hand activity already at 30 weeks of age (Locke et al., 1995), suggesting that gestures and speech go "hand-in-hand" from the earliest stages of cognitive development (McNeill, 1992; So et al., 2009).

Here we investigate whether gestures also convey some information about the prosodic structure of spoken language. We test whether prosody, an essential aspect of language, is also detected in gestures. In the auditory modality, prosody is characterized by changes in duration, intensity and pitch (for an overview see Cutler et al., 1997; Warren, 1999; Speer and Blodgett, 2006; Langus et al., 2012). Speakers can intentionally manipulate these acoustic cues to convey information about their states of mind (e.g., irony or sarcasm), to define the type of speech act they are making (e.g., a question or an assertion), and to highlight certain elements over others (e.g., by contrasting them). Importantly, prosody also conveys information about the structure of language. Because the grammatical structure of human language is automatically mapped onto prosodic structure during speech production (Langus et al., 2012), the prosody of spoken language

also signals the grammatical structure (Nespor and Vogel[1], 1986, 2007). Though prosody offers cues to different aspects of grammar, here we concentrate on the role of prosody in conveying information about syntactic structure.

It has been observed that prosodic cues are the most reliable cues for segmenting continuous speech cross-linguistically (Cutler et al., 1997). Adult listeners can use these cues to constrain lexical access (Christophe et al., 2004), to locate major syntactic boundaries in speech (Speer et al., 2011), and to determine how these relate to each other in sentences (Fernald and McRoberts, 1995; Langus et al., 2012). This is best seen in cases where listeners can disambiguate sentences that have more than one meaning (e.g., [bad] [boys and girls] vs. [bad boys] [and girls]) by relying on prosody alone (Lehiste et al., 1976; Nespor and Vogel, 1986, 2007; Price et al., 1991). Manipulations of the prosodic structure influence how listeners interpret syntactically ambiguous utterances (Lehiste, 1973; Lehiste et al., 1976; Cooper and Paccia-Cooper, 1980; Beach, 1991; Price et al., 1991; Carlson et al., 2001; see Cutler et al., 1997). These effects of prosody emerge quickly during online sentence comprehension, suggesting that they involve a robust property of the human parser (Marslen-Wilson et al., 1992; Warren et al., 1995; Nagel et al., 1996; Pynte and Prieur, 1996; Kjelgaard and Speer, 1999; Snedeker and Trueswell, 2003; Weber et al., 2006). Naive speakers systematically vary their prosody depending on the syntactic structure of sentences and naive listeners can use this variation to disambiguate utterances that—though containing the same sequence of words—differ in that they are mapped from sentences with different syntactic structures (Nespor and Vogel, 1986, 2007; Snedeker and Trueswell, 2003; Kraljic and Brennan, 2005; Schafer et al., 2005). These studies indicate that users of spoken language share implicit knowledge about the relationship between prosody and syntax and that they can use both during speech production and comprehension. To account for the syntax-prosody mapping, Nespor and Vogel (1986, 2007) have proposed a hierarchy that at the phrasal level contains—among other constituents—the Phonological Phrase (PP) and the Intonational Phrase (IP). These constituents are signaled in different ways: besides being signaled through external sandhi rules that are bound to a specific constituent, the PP right edge is signaled through final lengthening, and the IP level is signaled through pitch resetting at the left edge and through final lengthening at the right edge.

Here we ask whether prosody could also be perceived visually in the spontaneous gestures that accompany speech. In English and Italian, specific hand gestures ending with an abrupt stop, called "beats" (i.e., McNeill, 1992), are temporally related to pitch accents in speech production (Yasinnik et al., 2004; Esposito et al., 2007; Krahmer and Swerts, 2007). Also in sign languages, prosodic cues are not only conveyed through facial expressions, but also through hand and body movements (Nespor and Sandler, 1999; Wilbur, 1999; Sandler, 2011; Dachkovsky et al., 2013). A model developed on the basis of Israeli Signed Language showed that body positions align with rhythmic manual features of the signing stream to mark prosodic constituents' boundaries at different levels of the prosodic hierarchy (Nespor and Sandler, 1999; Sandler, 1999, 2005, 2011). More recently, Sandler (2012) proposed that many actions of the body in sign languages—that she calls "dedicated gestures"—perform linguistic functions and contribute to prosodic structure.

Do people perceive prosody and co-speech gestures as a coherent unit in everyday interactions? There is some evidence that both adults and infants match the global head and facial movements of the speaker with speech sounds (Graf et al., 2002; Munhall et al., 2004; Blossom and Morgan, 2006; Guellaï et al., 2011). However, it is unknown whether visual prosodic cues that accompany speech, but are not directly triggered by the movements of the vocal tract, are actually used to process the structure of the speech signal. Here we ask whether prosody can be perceived in the spontaneous gestures of a speaker (Experiment 1), and if listeners can use gestures to disambiguate sentences with the same sequence of words mapped onto different speech utterances that have two alternative meanings (Experiment 2). To investigate which prosodic cues participants rely on in disambiguating these sentences, we constructed sentences where disambiguation could be either due to IP or to PP boundaries. This enabled us to test whether the prosodic hierarchy is discernable from gestures alone.

EXPERIMENT 1

In this first experiment, we explored whether gestures carry prosodic information. We tested Italian-speaking participants in their ability to discriminate audio-visual presentations of low-pass filtered Italian utterances where the gestures either matched or mismatched the auditory stimuli (Singer and Goldin-Meadow, 2005). While low-pass filtering renders speech unintelligible, it preserves the prosody of the acoustic signal (Knoll et al., 2009). This guaranteed that only prosodic information was available to the listeners.

METHODS

Participants

We recruited 20 native speakers of Italian (15 females and 5 males, mean age 24 ± 5) from the subject pool of SISSA—International School of Advanced Studies (Trieste, Italy). Participants reported no auditory, vision, or language related problems. They received monetary compensation.

Stimuli

We used sentences that contain the same sequence of words and that can be disambiguated using prosodic cues at one of two different levels of the prosodic hierarchy. The disambiguation could take place at the IP level—the higher of these two constituents, coextensive with intonational contours—signaled through pitch resetting and final lengthening (Nespor and Vogel, 1986, 2007). For example, in Italian, *Quando Giacomo chiama suo fratello è sempre felice* is ambiguous because depending on the IP boundary *è sempre felice ((he) is always happy)* could refer to either *Giacomo* or *suo fratello* (*his brother*): (1) [Quando Giacomo chiama]$_{IP}$ [suo fratello è sempre felice]$_{IP}$ (*When Giacomo calls him his brother is*

[1]Though recursive prosodic phrasal constituents have been proposed at the level of the Intonational Phrase (Ladd, 1986) we rely on the more standardly accepted view that phrasal prosody has no recursive constituents (Selkirk, 1984; Nespor and Vogel, 1986, 2007).

always happy); or (2) [Quando Giacomo chiama suo fratello]$_{IP}$ [è sempre felice]$_{IP}$ (*When Giacomo calls his brother he is always happy)*.

Alternatively, the disambiguation could take place at the PP level where phrase boundaries are signaled through final lengthening. The PP extends from the left edge of a phrase to the right edge of its head in head-complement languages (e.g., Italian and English); and from the left edge of a head to the right edge of its phrase in complement-head languages (e.g., Japanese and Turkish) (Nespor and Vogel, 1986, 2007). An example of a phrase with two possible meanings is *mappe di città vecchie* that is ambiguous in Italian because depending on the location of the PP boundaries, the adjective *vecchie* (*old*) could refer to either *città* (towns) or *mappe* (maps): (1) [mappe di città]$_{PP}$ [vecchie]$_{PP}$ (old *maps of towns*); or (2) [mappe]$_{PP}$ [di città vecchie]$_{PP}$ (*maps of old towns*) (for more details see the list of the sentences ambiguous at the IP and PP levels used in Experiments 1 and 2 in **Table 1**). The presentation of the two types of sentences—those ambiguous at the IP level and those ambiguous at the PP level—was randomized across subjects.

We video recorded two native speakers of Italian—a male and a female—uttering ten different ambiguous Italian sentences (see **Table 1**). The speakers were unaware of the purpose or the specifics of the experiments. The speakers were asked to convey to an Italian listener the different meanings of the sentences using spontaneous gestures in the most natural way possible. They were video recorded under experimental conditions (i.e., not in natural setting) uttering the different sentences presented in **Table 1** with each of their two different meanings. The co-speech gestures produced contained both iconic gestures (i.e., gestures expressing some aspects of the lexical content) and beats ones (i.e., gestures linked to some prosodic aspects of the utterance) gestures (see Kendon, 1994 for a review; McNeill, 1992). The videos of the speakers were framed so that only the top of their body, from their shoulders to their waist, was visible (see **Movies S1**, **S2**). Thus, the mouth—i.e., the verbal articulation of the sentences—was not visible. Two categories of videos were created from these recordings using Sony Vegas 9.0 software. One category corresponded to the "matched videos" in which the speakers' gestures and their speech matched and the second category corresponded

Table 1 | Sentences ambiguous at the IP or PP level used in both Experiments with their prosodic parsing and their two possible meanings translated in English.

Sentences ambiguous at the Intonational Phrase level (IP)	Sentences ambiguous at the Phonological Phrase level (PP)
[[Alla conferenza]$_{PP}$ [Luciano]$_{PP}$ [ha parlato naturalmente]$_{PP}$]$_{IP}$ At the conference Luciano has talked in a natural way [[Alla conferenza]$_{PP}$ [Luciano]$_{PP}$ [ha parlato]$_{PP}$]$_{IP}$[[naturalmente]$_{PP}$]$_{IP}$ Of course Luciano talked at the conference	[[Come hai visto]$_{PP}$]$_{IP}$ [[la vecchia]$_{PP}$ [legge]$_{PP}$ [la regola]$_{PP}$]$_{IP}$ As you see the old woman reads the rule [[Come hai visto]$_{PP}$]$_{IP}$ [[la vecchia legge]$_{PP}$ [la regola]$_{PP}$]$_{IP}$ As you see the old law rules it
[[Come ti avevo detto]$_{PP}$]$_{IP}$ [quando Giorgio]$_{PP}$ [chiama]$_{PP}$]$_{IP}$ [[suo fratello]$_{PP}$ [è sempre nervoso]$_{PP}$]$_{IP}$ As I had told you when Giorgio calls his brother he is always happy [[Come ti avevo detto]$_{PP}$]$_{IP}$ [quando Giorgio]$_{PP}$ [chiama]$_{PP}$][suo fratello]$_{PP}$ [è sempre nervoso]$_{PP}$]$_{IP}$ As I had told you when Giorgio calls his brother is always happy	[[Come sicuramente hai visto]$_{PP}$]$_{IP}$ [la vecchia]$_{PP}$ [sbarra]$_{PP}$ [la porta]$_{PP}$]$_{IP}$ As you for sure have seen the old lady blocks the door [[Come sicuramente hai visto]$_{PP}$]$_{IP}$ [[la vecchia]$_{PP}$ [sbarra]$_{PP}$ [la porta]$_{PP}$]$_{IP}$ As you for sure have seen the old bar carries it
[[Come hai visto]$_{PP}$]$_{IP}$ [[quando Luca]$_{PP}$ [chiama]$_{PP}$ [il suo gatto]$_{PP}$]$_{IP}$ [è sempre felice]$_{PP}$] As you have seen when Luca calls his cat he is always happy [[Come hai visto]$_{PP}$]$_{IP}$ [[quando Luca]$_{PP}$ [chiama]$_{PP}$]$_{IP}$ [[il suo gatto]$_{PP}$ [è sempre felice]$_{PP}$]$_{IP}$ As you have seen when Luca calls his cat is always happy	[[Come ti avevo detto]$_{PP}$]$_{IP}$ [[quando Luca]$_{PP}$ [legge Dante]$_{PP}$ [è felice]$_{PP}$]$_{IP}$ As I had told you when Luca reads Dante he is happy [[Come ti avevo detto]$_{PP}$]$_{IP}$ [[quando Luca]$_{PP}$ [legge]$_{PP}$ [Dante]$_{PP}$ [è felice]$_{PP}$]$_{IP}$ As I had told you when Luca reads Dante is happy
[[Come ti avevo detto]$_{PP}$]$_{IP}$ [[se Giacomo]$_{PP}$ [scrive bene]$_{PP}$ [è felice]$_{PP}$]$_{IP}$ As I had told you if Giacomo writes well he is happy [[Come ti avevo detto]$_{PP}$]$_{IP}$] [[se Giacomo]$_{PP}$ [scrive]$_{PP}$]$_{IP}$] [[Bene]$_{PP}$ [è felice]$_{PP}$]$_{IP}$ As I had told you if Giacomo calls Bene is happy	[[Sanno]$_{PP}$ [tutti]$_{PP}$ [che canta solo]$_{PP}$ [se è felice]$_{PP}$]$_{IP}$ Everybody knows that he sings alone if he is happy [[Sanno]$_{PP}$ [tutti]$_{PP}$ [che canta]$_{PP}$ [solo]$_{PP}$ [se è felice]$_{PP}$]$_{IP}$ Everybody knows that he sings only if he is happy
[[Sai]$_{PP}$ [che parla]$_{PP}$ [molte lingue]$_{PP}$ [naturalmente]$_{PP}$]$_{IP}$ You know that he speaks many languages in a natural way [[Sai]$_{PP}$ [che parla]$_{PP}$ [molte lingue]$_{PP}$]$_{IP}$ [[naturalmente]$_{PP}$]$_{IP}$] You of course know that he speaks many languages	
[Come ti avevo detto]$_{PP}$]$_{IP}$ [salta]$_{PP}$ [il muro]$_{PP}$ [più alto]$_{PP}$[naturalmente]$_{PP}$]$_{IP}$ As I had told you s/he jumps over the tallest wall in a natural way [Come ti avevo detto]$_{PP}$]$_{IP}$ [salta]$_{PP}$ [il muro]$_{PP}$ [più alto]$_{PP}$]$_{IP}$[[naturalmente]$_{PP}$]$_{IP}$ As I had told you of course s/he jumps over the tallest wall	

to the "mismatched videos" in which the gestures were associated with the speech sound of the same sequence of words, but with the alternative meaning. To do so, we edited the original recordings and switched the acoustic and visual stimuli. This manipulation was not perceived by the participants as reported in the debriefing session. Then the gestures signaled the opposite meaning of that is signaled by the sentence for this condition. A total of 80 videos were created (each of the sentences was uttered twice). We ensured that, in the mismatched audio-visual presentations, the left and the right edges of the gesture sequences were aligned with the left and the right edges of the utterances (see **Figure 1**). This is an important point as in sign languages manual alignment with the signing stream is quite strict (Nespor and Sandler, 1999; Sandler, 2012) and co-speech gestures in general are tightly temporally linked to speech (McNeill et al., 2000). To remove the intelligibility of speech but to preserve prosodic information, the speech sounds were low-pass filtered using Praat software with the Haan band filter (0–400 Hz). As a result it was not possible to detect from speech which of the two meanings of a sentence was intended, as reported by the participants at the end of the experiment. The resulting stimuli had the same loudness of 70 dB.

PROCEDURE

Participants were tested in a soundproof room and the stimuli were presented through headphones. They were instructed to watch the videos and answer—by pressing a key on a keyboard—whether what they saw matched or mismatched what they heard (i.e., [S] = yes or [N] = no). A final debriefing (i.e., we explained the goals of the study) ensured that none of the participants understood the meaning of the sentences.

RESULTS AND DISCUSSION

The results show that participants correctly identified the videos in which hand gestures and speech matched [$M = 81.9$, $SD = 11.03$: t-test against chance with equal variance not assumed $t_{(19)} = 12.93, p < 0.0001$] and those in which they did not match [$M = 69.3$, $SD = 10.17$; $t_{(19)} = 8.41$, $p < 0.0001$]. A repeated measure ANOVA with condition (Match, Mismatch) and type of prosodic contour (IP and PP) was performed on the mean percentage. The ANOVA only revealed a significant main effect for condition [$F_{(1, 19)} = 12.81, p = 0.002, \acute{\eta}^2 = 0.4$], but neither for type of prosodic contour [$F_{(1, 19)} = 1.20$, $p = 0.287$, $\acute{\eta}^2 = 0.06$] nor for an interaction of type and condition [$F_{(1, 19)} = 3.52$, $p = 0.076$, $\acute{\eta}^2 = 0.16$]. Participants answered correctly more often in the matching condition, and there are more errors for the mismatching one. In other words, they are more likely to incorrectly accept a mismatching video than to reject a matching one. A possible interpretation for this asymmetric results is that participants may detect some incoherences in the mismatching videos and these could lead them to a certain degree of uncertainty in their answers. To sum up, the results show that adult listeners detect the congruency between hand gestures and the acoustic speech signal even when only the prosodic cues are preserved in the acoustic signal (see **Figure 2**). The spontaneous gestures that accompany speech must therefore be aligned with the speech signal, suggesting a tight link between the motor-programs responsible for producing both speech and the spontaneous gestures that accompany it. The results of Experiment 1 thus show that adult listeners are sensitive to the temporal alignment of speech and the gestures that speakers spontaneously produce when they speak. In the next Experiment we asked whether the gestures that accompany speech

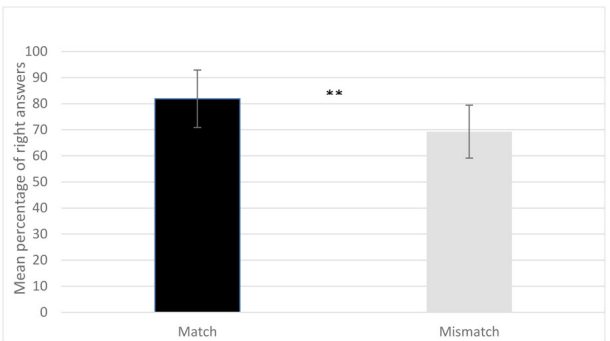

FIGURE 2 | Mean percentage of right answers in the match and mismatch conditions of Experiment 1. Participants' mean percentage of right answers is significantly higher in the matching condition than in the mismatching one (**$p < 0.0001$). Errors bars represent the standard deviation.

FIGURE 1 | Examples of the stimuli used in both Experiments (i.e., with speech being filtered for Experiment 1). Here the sentence is "Come hai visto quando Luca chiama il suo gatto è sempre felice." Two meanings are possible: "As you have seen when Luca calls his cat is always happy" (meaning 1) vs. "As you have seen when Luca calls his cat he is always happy" (meaning 2). On the left, this is the matched version (i.e., the audio and the visual inputs match) whereas on the right this is the mismatched version (i.e., the audio of meaning 1 is aligned with the visual input of meaning 2). The left and right edges of gesture sequences and those of utterances were aligned.

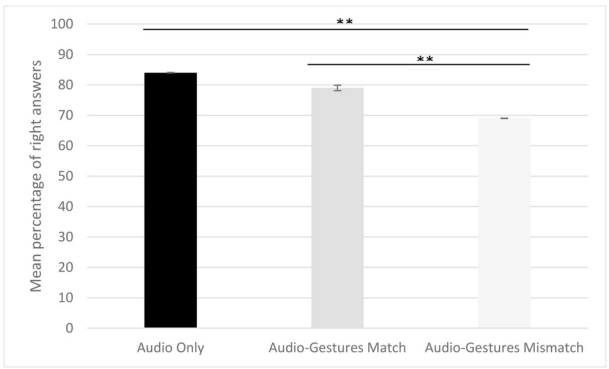

FIGURE 3 | Mean percentage of right answers in the audio only, the match and mismatch conditions of Experiment 2. Participants' mean percentage of right answers is higher in the audio and matching conditions than in the mismatching one (**$p < 0.0001$). Errors bars represent the standard deviation.

have any effect on adult listeners' understanding of ambiguous sentences.

EXPERIMENT 2

In sign languages, a good deal of prosodic information is conveyed by gestures of different parts of the face and body (Sandler, 2012). This information alone can distinguish coordinate from subordinate sentences and declarative sentences from questions (Pfau and Quer, 2010; Dachkovsky et al., 2013). This may suggest that in spoken languages too, listeners can actively use gestures accompanying speech for perceiving, processing and also understanding speech. For example, if gestures are carrying prosodic information about the grammatical structure of the speech signal, it should be easier for listeners to disambiguate a sentence that can have two different meanings when the gestures accompanying speech are visible and match the audible utterance. Experiment 2 was designed to test this hypothesis. We presented to Italian-speaking adults potentially ambiguous Italian sentences in which the audio-visual information was either matched or mismatched.

METHODS

Participants

We recruited 20 native speakers of Italian (9 females and 11 males, mean age 23 ± 3) from the subject pool of SISSA—International School of Advanced Studies (Trieste, Italy). Participants reported no auditory, vision, or language related problems. They received monetary compensation.

Stimuli

The same videos of the speakers recorded for Experiment 1 were used. However, for Experiment 2, the speech sound was not low-pass filtered (see **Movies S3**, **S4**). We added also audio-only samples of the sentences as a control condition. Thus, there were three categories of stimuli for Experiment 2: auditory only, auditory with matched gestures and auditory with mismatched gestures. For each of the categories, there were 10 different sentences (i.e., the same sentences as in Experiment 1) that could have two different meanings, uttered by a male and a female speaker. Thus, a total of 120 stimuli were created. We ensured that the left and right edges of gesture sequences and those of utterances were aligned. Speech sounds for all the stimuli had the same loudness of 70 dB.

PROCEDURE

Participants were tested in a soundproof room with headphones. They were instructed to both listen to and to watch the stimuli. After each presentation, a question appeared on the screen regarding the meaning of the sentence they had just perceived. For example, after "Quando Giacomo chiama suo fratello è sempre felice" (When—Giacomo—calls—his—brother—is—always - happy) either the question "Giacomo è felice?" *(Is Giacomo happy?)*, or the question "Suo fratello è felice?" *(Is his brother happy?)* appeared. Participants had to answer, by clicking on a keyboard, if the answer to the question was *yes* or *no*. In each of the three within-subject conditions (audio only, audio and gestures match, audio and gestures mismatch) participants saw 5 of the 10 sentences (total 10 different meanings) so that each meaning was paired with a "yes" question ("yes" = hit/"no" = miss) and a "no" question ("yes" = correct rejection/"no" = false alarm). Each participant heard the same sentence produced by the female and the male speaker resulting in a total of 120 trials.

RESULTS

First, comparisons against chance indicated that participants' overall accuracy of the presented stimuli was significantly above chance (see **Figure 3**) [Audio condition: $M = 84.1$, $SD = 9.2$: t-test against chance with equal variance not assumed $t_{(19)} = 24.7$, $p < 0.0001$; Match condition: $M = 79$, $SD = 8.8$, $t_{(19)} = 23.5$, $p < 0.0001$; Mismatch condition: $M = 69.1$, $SD = 5.2$, $t_{(19)} = 31$, $p < 0.0001$]. In order to determine participants' performance in each of the three conditions we calculated the F-score (2*accuracy*completeness)/(accuracy+completeness): the harmonic mean of Accuracy [#hits/(#hits+#false alarms)] and Completeness (#hits/(#hits+#misses)). We ran a repeated measures ANOVA with Condition (Audio Only, Audio-Gesture Match, Audio-Gesture Mismatch) and Type of Prosodic Contour (IP and PP) as within-subject factors. We found a significant main effect for condition [$F_{(2,\ 18)} = 20.1$, $p = 0.0001$, $\acute{\eta}^2 = 0.7$], a marginally significant effect for Type [$F_{(1,\ 19)} = 4.226$, $p = 0.054$, $\acute{\eta}^2 = 0.18$] and a significant interaction of Type and Condition [$F_{(2,\ 18)} = 14.624$, $p < 0.0001$, $\acute{\eta}^2 = 0.6$]. Paired sample t-tests used for *post-hoc* comparisons (Bonferroni correction $p < 0.0083$) revealed a significant difference between Audio Only ($M = 84.1$, $SD = 9.2$) and Audio-Gestures Mismatch ($M = 69.1$, $SD = 5.2$) conditions [$t_{(19)} = 6.78$, $p < 0.0001$], and between Audio-Gesture Match ($M = 79$, $SD = 8.8$) and Audio-Gesture Mismatch conditions [$t_{(19)} = 4.67$, $p < 0.0001$], but not between Audio only and Audio-Gesture Match conditions [$t_{(19)} = 1.40$, $p = 0.178$]. While the type of the prosodic contour did not affect participants' performance in the Audio only condition [$M_{IP} = 87$, $SD_{IP} = 10$; $M_{PP} = 79$, $SD_{PP} = 13$: $t_{(19)} = 2.408$, $p = 0.026$], participants performed significantly better on sentences disambiguated with PP than on sentences disambiguated with IP boundaries in Audio-Gesture Match [$M_{IP} = 75$,

$SD_{IP} = 11$; $M_{PP} = 85$, $SD_{PP} = 12$: $t_{(19)} = -3.105$, $p = 0.006$] and Audio-Gesture mismatch [$M_{IP} = 64$, $SD_{IP} = 8$; $M_{PP} = 70$, $SD_{PP} = 10$: $t_{(19)} = -3.376$, $p = 0.003$] conditions. First, these results show that matching gestures do not lead to a better comprehension than audio alone, while mismatching gestures hinder comprehension. Second, when the prosody of gestures mismatched that of speech, participants could not ignore the mismatch in their effort to disambiguate sentences. Interestingly, while on the whole, perceiving speech with and without gestures did not appear to influence sentence comprehension as scores are above chance level, participants have more difficulties to disambiguate sentences with IP than with PP boundaries both in the gestures matched and in the gestures mismatched conditions.

GENERAL DISCUSSION

Our findings show that when presented with acoustic linguistic stimuli that contain only prosodic information (i.e., low-pass filtered speech), participants are highly proficient in detecting whether speech sounds and gestures match. The prosodic information of spoken language must therefore be tightly connected to gestures in speech production that are exploited in speech perception. The syntactic structure and the meaning of utterances appear thus not to be necessary for the perceiver to align gestures and prosody. Additionally, participants could also use co-speech gestures in their comprehension of potentially ambiguous sentences, i.e., sentences with the same sequence of words, thus totally ambiguous in their written form, but with different prosodic structures. The disambiguation of these sentences could be triggered either by the PP or by the IP division into constituents. Our results show that matching gestures do not lead to a better comprehension than audio alone, while mismatching gestures led participants to choose significantly more the meaning signaled by gestures. Therefore, gestures are used in interpreting the meaning of ambiguous sentences. Interestingly, in the presence of gestures, participants have more difficulties to disambiguate sentences with IP than with PP boundaries in both conditions. These results suggest that the presence of gestures impairs performances when auditory cues are stronger. For example, it is possible that PPs are less marked by auditory cues than the IPs and therefore gestures might give additional information in this case. It seems also important here to point out the fact that in the present study what we call mismatch videos are videos in which the audio file of one meaning of a sentence is presented with the image video of the alternative meaning of the same sentence. Therefore, this manipulation of stimuli could have led to a possible artifact in the participants' performances. Though this possibility cannot be excluded entirely, we believe it is unlikely. At the end of the test session, we asked participants whether they had noticed the mismatching manipulation. None of the participants tested reported any perception of a manipulation. Thus, when they had the two categories of sentences, matched and mismatched, they did not detect that they were different because one was manipulated and not the other.

As opposed to the visual perception of speech in the speakers' face, where the movements of the mouth, the lips, but also the eyebrows (Krahmer and Swerts, 2004) are unavoidable in the production of spoken language, the gestures that accompany speech belong to a different category that is avoidable in speech production. Even though mismatching gestures decrease the intelligibility of spoken language, the addition of matching gestures does not appear to give an advantage over speech perception in the auditory modality alone. We are, in fact, able to understand the meaning of sentences when talking on the phone, or if our interlocutor is for other reasons invisible. Our results, however, suggest that the prosody of language extends from the auditory to the visual modality in speech perception.

This link between speech and gestures is congruent with neuropsychological evidence for a strong correlation between the severity of aphasia and the severity of impairment in gesturing (Cocks et al., 2013). While further studies are clearly needed to identify the specific aspects of spontaneous gestures that are coordinated with speech acts, our results demonstrate that part of speech perception includes the anticipation that bodily behaviors, such as gestures, be coordinated with speech acts. Prosodic Phonology thus appears—at least in part—not to be a property exclusive to oral language. In fact, it has abundantly been shown to characterize also sign languages where it has an influence on all body movements (Nespor and Sandler, 1999; Wilbur, 1999; Sandler, 2011, 2012). It is also—at least in part—not specific to language. Previous findings have shown that part of prosody, i.e., rhythmic alternation as defined by the Iambic—Trochaic Law (Bolton, 1894; Nespor et al., 2008; Bion et al., 2011) characterizes also the grouping of non-linguistic visual sequences (Peña et al., 2011). Thus, language is a multimodal experience and some of its characteristics are domain-general rather than domain-specific.

ACKNOWLEDGMENT

The present research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement n° 269502 (PASCAL), and the Fyssen Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00700/abstract

Movie S1 | One sentence low-pass filtered speech, match condition (Experiment 1).

Movie S2 | Same sentence low-pass filtered speech, mismatch condition (Experiment 1).

Movie S3 | One sentence, match condition (Experiment 2).

Movie S4 | Same video, mismatch condition (Experiment 2).

REFERENCES

Alibali, M. W., Kita, S., and Young, A. J. (2000). Gesture and the process of speech production: we think, therefore we gesture. *Lang. Cogn. Process.* 15, 593–613. doi: 10.1080/016909600750040571

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: evidence for cue trading relations. *J. Mem. Lang.* 30, 644–663. doi: 10.1016/0749-596X(91)90030-N

Bernstein, L. E., Eberhardt, S. P., and Demorest, M. E. (1998). Single-channel vibrotactile supplements to visual perception of intonation and stress. *J. Acoust. Soc. Am.* 8, 397–405.

Bion, R. H., Benavides, S., and Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' memory for speech sequences. *Lang. Speech* 54, 123–140. doi: 10.1177/0023830910388018

Blossom, M., and Morgan, J. L. (2006). "Does the face say what the mouth says? A study of infants' sensitivity to visual prosody," in *Proceedings of the 30th Annual Boston University Conference on Language Development*, eds D. Bamman, T. Magnitskaia, and C. Zaller (Somerville, MA: Cascadilla Press).

Bolton, T. (1894). Rhythm. *Am. J. Psychol.* 6, 145–238. doi: 10.2307/1410948

Carlson, K., Clifton, C., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *J. Mem. Lang.* 45, 58–81. doi: 10.1006/jmla.2000.2762

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *J. Mem. Lang.* 51, 523–547. doi: 10.1016/j.jml.2004.07.001

Chu, M., and Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *J. Exp. Psychol. Gen.* 140, 102–115. doi: 10.1037/a0021790

Cocks, N., Dipper, L., Pritchard, M., and Morgan, G. (2013). The impact of impaired semantic knowledge on spontaneous iconic gesture production. *Aphasiology* 27, 1050–1069. doi: 10.1080/02687038.2013.770816

Cook, S. M., and Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *J. Cogn. Dev.* 7, 211–232. doi: 10.1207/s15327647jcd0702_4

Cooper, W. E., and Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674283947

Corballis, M. C. (2002). *From Hand to Mouth: the Origins of Language*. Princeton; Oxford: Princeton University Press.

Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40, 141–201.

Dachkovsky, S., Healy, C., and Sandler, W. (2013). Visual intonation in two sign languages *Phonoly* 30, 211–252. doi: 10.1017/S0952675713000122

de Ruiter, J. P., Bangerter, A., and Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: investigating the tradeoff hypothesis. *Top. Cogn. Sci.* 4, 232–248. doi: 10.1111/j.1756-8765.2012.01183.x

Esposito, A., Esposito, D., Refice, M., Savino, M., and Shattuck-Hufnagel, S. (2007). "A preliminary investigationof the relationship between gestures and prosody in Italian," in *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, eds A. Esposito, M. Bratanic, E. Keller, and M. Marinaro (Amsterdam: IOS Press), 45–57.

Fernald, A., and McRoberts, G. W. (1995). "Prosodic bootstrapping: a critical analysis of the argument and the evidence," in *Signal to Syntax: Bootstrapping from Speech to Syntax in Early Acquisition,* eds J. L. Morgan and C. Demuth (Hillsdale, NJ: Erlbaum Associates), 365–387.

Graf, P. H., Cosatto, E., Strom, V., and Huang, F. J. (2002). "Visual prosody: Facial movements accompanying speech," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition* (Washington, DC). doi: 10.1109/AFGR.2002.1004186

Guellaï, B., Coulon, M., and Streri, A. (2011). The role of motion and speech in face recognition at birth. *Vis. Cogn.* 19, 1212–1233. doi: 10.1080/13506285.2011.620578

Iverson, J. M., and Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature* 396, 228.

Kelly, S., and Barr, D. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *J. Mem. Lang.* 40, 577–592.

Kendon, A. (1983). Gesture. *J. Vis. Verb. Lang.* 3, 21–36.

Kendon, A. (1994). Do gestures communicate? *Res. Lang. Soc. Inter.* 27, 175–200.

Kjelgaard, M. M., and Speer, S. R. (1999). Prosodic facilitaion and interference in the resolution of temporary syntactic closure ambiguity. *J. Mem. Lang.* 40, 153–194. doi: 10.1006/jmla.1998.2620

Knoll, M. A., Uther, M., and Costall, A. (2009). Effects of low-pass filtering on the judgement of emotional affect using speech directed to infants, adults and foreigners. *Speech Commun.* 51, 210–216. doi: 10.1016/j.specom.2008.08.001

Krahmer, E. J., and Swerts, M. (2004). "More about brows: a cross-linguistic analysis-by-synthesis study," in *From Brows to Trust: Evaluating Embodied Conversational Agents,* eds C. Pelachaud and Z. S. Ruttkay (Antwerp: Kluwer Academic Publishers), 191–216.

Krahmer, E. J., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005

Kraljic, T., and Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: for the speaker or for the addressee? *Cogn. Psychol.* 50, 194–231. doi: 10.1016/j.cogpsych.2004.08.002

Ladd, D. R. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonol. Yearbook* 3, 311–340. doi: 10.1017/S0952675700000671

Langus, A., Marchetto, E., Bion, R. A., and Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *J. Mem. Lang.* 66, 285–306. doi: 10.1016/j.jml.2011.09.004

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa* 7, 102–122.

Lehiste, I., Olive, J. P., and Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *J. Acoust. Soc. Am.* 60, 1199–1202. doi: 10.1121/1.381180

Locke, J. L., Bekken, K. E., McMinn-Larsen, L., and Wein, D. (1995). Emergent control of manual and vocal-motor activity in relation to the development of speech. *Brain Lang.* 51, 498–508. doi: 10.1006/brln.1995.1073

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., and Lee, C. S. (1992). Prosodic effects in minimal attachment. *Q. J. Exp. Psychol.* 45, 73–87. doi: 10.1080/14640749208401316

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

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

McNeill, N. M., Alibali, M. W., and Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: now they need it, now they don't. *J. Nonverbal Behav.* 24, 131–150. doi: 10.1023/A:1006657929803

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and VatikiotisBateson, E. (2004).Visual prosody and speech intelligibility. *Psychol. Sci.* 15, 133–137. doi: 10.1111/j.0963-7214.2004.01502010.x

Nagel, H. N., Shapiro, L. P., Tuller, B., and Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *J. Psycholinguist. Res.* 25, 319–344. doi: 10.1007/BF01708576

Nespor, M., M., Shukla, R., van de Vijver, C., Avesani, H., and Schraudolf, and, C., Donati (2008). Different phrasal prominence realization in VO and OV languages. *Lé&L* 7, 1–28.

Nespor, M., and Sandler, W. (1999). Prosody in israeli sign language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris, 327.

Nespor, M., and Vogel, I. (2007). *Prosodic Phonology*. 1st Edn. 1986. Berlin: Mouton De Gruyter.

Peña, M., Bion, R. H., and Nespor, M. (2011). How modality specific is the iambic-trochaic law? Evidence from vision. *J. Exp. Psychol. Lang. Mem. Cogn.* 37, 1199–1208. doi: 10.1037/a0023944

Pfau, R., and Quer, J. (2010). "Nonmanuals: their grammatical and prosodic roles," in *Sign Languages,* ed D. Brentani (Cambridge: Cambridge University Press), 381–402.

Ping, R., and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about non-present objects. *Cogn. Sci.* 34, 602–619. doi: 10.1111/j.1551-6709.2010.01102.x

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.* 90, 2956–2970. doi: 10.1121/1.401770

Pynte, J., and Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Lang. Cogn. Process.* 11, 165–192. doi: 10.1080/016909696387259

Rimé, B., and Shiaratura, L. (1991). "Gesture and speech," in *Fundamentals of Non-Verbal Behaviour,* eds R. S. Feldman and B. Rime (Cambridge: Cambridge University Press), 239–281.

Sandler, W. (1999). Prosody in two natural language modalities. *Lang. Speech* 42, 127–142. doi: 10.1177/00238309990420020101

Sandler, W. (2005). Prosodic constituency and intonation in sign language. *Linguistische Berichte* 13, 59–86.

Sandler, W. (2011). "The phonology of movement in sign language," in *Blackwell Companion to Phonology,* eds M. van Oostendorp, C. Ewen, K. Rice, and E, Hume (Oxford: Wiley-Blackwell), 577–603.

Sandler, W. (2012). Dedicated gestures the emergence of sign language. *Gesture* 12, 265–307. doi: 10.1075/gest.12.3.01san

Schafer, A., Speer, S., and Warren, P. (2005). "Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task,"

in *Approaches to Studying World Situated Language Use,* eds M. Tanenhaus and J. Trueswell (Cambridge: MIT Press).

Selkirk, E. O. (1984). *Phonology and Syntax: the Relation Between Sound and Structure*. Cambridge: MIT Press.

Singer, M. A., and Goldin-Meadow, S. (2005). Children learn when their teachers gestures and speech differ. *Psychol. Sci.* 16, 85–89. doi: 10.1111/j.0956-7976.2005.00786.x

Snedeker, J., and Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *J. Mem. Lang.* 48, 103–130. doi: 10.1016/S0749-596X(02)00519-3

So, W. C., Kita, S., and Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cogn. Sci.* 33, 115–125. doi: 10.1111/j.1551-6709.2008.01006.x

Speer, S. R., and Blodgett, A. (2006). "Prosody," in *Handbook of Psycholinguistics, 2nd Edn.,* eds M. Traxler and M. A. Gernsbacher (Amsterdam: Elsevier), 505–537.

Speer, S. R., Warren, P., and Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Lab. Phonol.* 2, 35–98. doi: 10.1515/labphon.2011.002

Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309

Warren, P. (1999). "Prosody and sentence processing," in *Language Processing,* eds S. Garrod and M. Pickering (Hove: Psychology Press), 155–188.

Warren, P., Grabe, E., and Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Lang. Cogn. Process.* 10, 457–486. doi: 10.1080/01690969508407112

Weber, A., Grice, M., and Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: a study of nticipatory eye movements. *Cognition* 99, B63–B72. doi: 10.1016/j.cognition.2005.07.001

Wilbur, R. B. (1999). Stress in ASL: Empirical evidence and linguistic issues. *Lang. Speech* 42, 229–250. doi: 10.1177/00238309990420020501

Yasinnik, Y., Renwick, M., and Shattuck-Hufnagel, S. (2004). "The timing of speech-accompanying gestures with respect to prosody," in *From Sound to Sense: 50+ Years of Discoveries in Speech Communication, 11–13 June 2004,* (Cambridge, MA).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2014; accepted: 18 June 2014; published online: 07 July 2014.
*Citation: Guellaï B, Langus A and Nespor M (2014) Prosody in the hands of the speaker. Front. Psychol. **5**:700. doi: 10.3389/fpsyg.2014.00700*
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.