Coverbal Speech Gestures Do Not Impact Preschoolers' Ability to Use Prosodic Information to Constrain Parsing

Leticia Schiavon Kolberg, Elodie Charpentier, Alex de Carvalho

BUCLD 48 Proceedings
To be published in 2024 by Cascadilla Press
Rights forms signed by all authors

Coverbal Speech Gestures Do Not Impact Preschoolers' Ability to Use Prosodic Information to Constrain Parsing

Leticia Schiavon Kolberg, Elodie Charpentier, and Alex de Carvalho

1. Introduction

Prosodic information (e.g., syllable lengthening and intonation variations naturally produced in speech) has been shown to be an important source of information that children can use to parse the syntactic structure of sentences they hear, because prosodic boundaries in speech often coincide with syntactic boundaries (Nespor & Vogel, 1986; Morgan & Demuth, 1996). Recent studies show that young children can indeed rely on their perception of prosodic boundaries in speech to constrain parsing (e.g., de Carvalho, Dautriche & Christophe, 2016; de Carvalho, Kolberg, Trueswell & Christophe, 2022; Kolberg et al., 2021; Massicotte-Laforge & Shi, 2015; Snedeker & Yuan, 2008). For instance, de Carvalho, Dautriche & Christophe (2016) have shown that French preschoolers can rely on their perception of prosodic boundaries to constrain their interpretation of a noun-verb homophone such as *souri*, which can mean either the verb "to smile", in a sentence such as [*Tu vois?*][*le bébé sourit!*] ([Do you see?][The baby smiles!]) or the noun "mouse", in a sentence such as [*Tu vois le bébé souris?*] ([Do you see the baby mouse?]; the brackets indicate the different prosodic phrasings of the sentences). Participants listened to one of these two types of sentence while seeing two images side-by-side: one representing the noun interpretation of the homophone (e.g., a mouse) and another representing the verb interpretation (e.g., a baby smiling). Children who listened to the homophones inside sentences with noun prosody looked significantly longer towards the image representing the noun than the ones who listened to the sentences with verb prosody, suggesting that they were able to use prosodic boundary information to constrain their interpretation of the sentences.

Studies such as the one above suggest that young children understand the correlation between prosodic and syntactic phrases in speech, and can use

[1] * Leticia Schiavon Kolberg, Laboratoire de Psychologie du Développement et de l'Éducation de l'Enfant (LaPsyDÉ, Université Paris Cité - CNRS), leticia.kolberg@gmail.com*. Elodie Charpentier* and Alex de Carvalho, (LaPsyDÉ, Université Paris Cité - CNRS). *Joint first authorship. We thank all children, parents, schools and adults who participated in this study. This project has received financial support from the CNRS through the MITI interdisciplinary programs, and from Sciences Po – Université Paris Cité (Fonds Investissements d'Avenir), all awarded to Alex de Carvalho.

prosodic boundary information as a cue to syntactic structure during sentence processing. Other studies extended these findings to other languages (e.g., English and Brazilian Portuguese: de Carvalho & Kolberg et al., 2022; English: Snedeker & Yuan, 2008) and to more complex and less frequent structures (e.g., stripping (TP ellipsis) sentences: de Carvalho & Kolberg et al., 2022; Kolberg et al., 2021), suggesting that this ability is not restricted to French and/or to specific types of structures/ambiguities. Prosodic information is therefore assumed to be an important cue for syntactic acquisition and parsing in young children and perhaps universally.

However, acoustic information *per se* may not be the only cue to prosodic/syntactic structure that children and adults use when parsing sentences. Recent studies suggest that coverbal facial speech gestures, such as head nods and eyebrow movements naturally produced by speakers when uttering sentences, tend to align with prosodic boundaries in speech (e.g., de la Cruz-Pavía et al., 2020a; Esteve-Gibert & Prieto, 2013). For instance, in a sentence production task conducted in both English and Japanese (de la Cruz-Pavía et al., 2020a), naive adults were shown to coordinate eyebrow movements with their prosodic phrasing of elicited sentences. Participants were instructed to silently read sentences such as "In English, [behind mountains] is a phrase", where the phrase between brackets (the target phrase) varied across test trials, and then repeat them while looking at a camera, as if they were uttering the sentence to another person (i.e., either a child or an adult). Participants were shown to raise their eyebrows at the onset of the target phrase (e.g., at the beginning of "behind"), and lower them upon completing the last word of the phrase (e.g., "mountains"). These results show that adults used eyebrow movements to signal the boundaries between the most important prosodic unit in the sentence, namely, the target phrase that varied across test sentences.

If adults' facial coverbal speech gestures tend to align with prosodic phrasing in natural speech, children could potentially use this information as an additional cue for identifying prosodic and, consequently, syntactic boundaries. This possibility is supported by previous research demonstrating that oral speech perception is indeed multimodal, incorporating not only auditory cues but sometimes also additional visual information (e.g., Burnham & Dodd, 2004; McGurk & MacDonald, 1976; Rosenblum, Schmuckler, & Johnson, 1997). For instance, McGurk & MacDonald (1976) have shown that visual information can influence auditory perception of phonemes: when listening to a syllable such as /ba/, while seeing a speaker uttering a different syllable, such as /ga/, adults report perceiving /da/, a third syllable resulting from the fusion of the auditory and visual information. This phenomenon, known in the literature as the *McGurk effect*, has been demonstrated even in 5-month-old infants (Rosenblum, Schmuckler, & Johnson, 1997), showing that infants can integrate visual and auditory information in speech perception from an early age.

It remains to be shown however whether young children can benefit from facial coverbal speech gestures aligning with prosodic phrasing in natural speech to constrain their parsing. To investigate the potential use of these gestures as cues

to prosodic boundaries in speech, de la Cruz-Pavía et al. (2020b) conducted a task in which adult participants were asked to segment phrases in an artificial language. The results revealed that participants performed better in parsing unknown languages into phrase-like units when prosodic cues were accompanied by co-speech (facial) gestures aligning with prosodic phrasing, compared to when they listened to the same strings of syllables containing only prosodic cues without the gestures. This suggests that adult listeners can benefit from additional visual information to prosodic boundaries in speech, and can use it as a cue for finding phrase boundaries.

Despite evidence of infants and adults integrating auditory and visual information during language processing, and adults using coverbal speech gestures to guide parsing of an artificial language, it remains unclear whether coverbal speech gestures could impact listeners' ability to use prosody to constrain syntactic analysis of natural languages. To investigate this question, we tested whether children and adults' ability to use prosody to constrain parsing of sentences can be affected by the presence of both auditory and visual cues to prosodic boundaries, compared to solely auditory information.

We adapted a sentence completion task from de Carvalho, Dautriche & Christophe (2016). In their original task, children listened to the beginnings of sentences containing noun/verb homophones in French, such as *ferme*, which can mean either the verb "to close", in a sentence such as [*La petite*][***ferme** le coffre à jouets*] ([The little (girl)][**closes** the toy box]), or the noun "farm", in a sentence such as [*La petite **ferme***][*lui plaît beaucoup*] ([The little **farm**][pleases him a lot]). Children watched videos of a woman uttering one of these two types of sentence, but the audio/video was cut right after the target word and replaced with babble noise, such that the only information they had available to disambiguate the meaning of the homophone were the different prosodic phrasings at the beginning of the sentences until the offset of the ambiguous word. Sentences with noun prosody presented all three audible words within the same prosodic phrase (e.g., [*La petite ferme*]...), whereas in the sentences with verb prosody the same three words appeared in two different prosodic units (e.g., [*La petite*][*ferme*…) and the homophone was preceded by a prosodic boundary, coinciding with the boundary between the noun phrase (e.g., [*La petite*]) and the verb phrase (e.g., [*ferme*…]). Participants heard the beginnings of these sentences either in the noun prosody condition or in the verb prosody condition, and were then asked to complete the sentences with their own words. In their completions, children were shown to interpret the homophone more often as a noun (e.g., by completing "*la petite ferme…*" with a verb phrase, such as "*sera pour les enfants*" - "the little farm…will be for children") in the noun prosody condition than in the verb prosody condition, where they more often used the ambiguous word as a verb (e.g., saying something like "*La petite ferme… la fenêtre*" - "The little girl closes… the window").

As participants in de Carvalho et al. (2016) were exposed to a video of the speaker uttering the test sentences, they might have benefited from both visual and auditory cues to prosodic boundaries to constrain their syntactic analysis.

Indeed, although the speaker was simply instructed to utter the sentences while looking at the camera as if she was speaking to a child, she naturally produced facial coverbal speech gestures (eyebrow movements and head nods) around prosodic boundaries. However, this possibility was not explored in their study, as they did not assess participants' performance across different modalities (e.g., comparing the ability to use prosody for parsing with both visual and auditory cues versus solely with auditory cues).

To assess whether the presence of visual cues enhanced children's ability to use prosodic information to constrain parsing in de Carvalho et al.'s study, we conducted this task with two groups of participants: one group was presented with the original task, with both visual and auditory information to prosodic boundaries, while the other group engaged in a modified version of this task with the same auditory stimulus but without the visual information presented through the speaker's videos (i.e., we replaced the video of the speaker by a picture of a radio). If children exposed to sentences in both visual and auditory modalities perform better than the ones exposed to the new audio-only condition, this would suggest that they benefit from simultaneous visual and auditory cues to prosodic boundaries to guide their parsing.

2. Experiment 1

2.1. Participants

Forty-eight 4-to-6-year-old native French-speaking children (M_{age} = 5.7 years old, range 4.4 to 6.3 years, 24 girls) and forty-eight native French-speaking adults (M_{age} = 21.5 years old, range 17 to 56 years, 44 women) were included in the final sample. Half of the participants were assigned to the audio-visual condition, and the other half was assigned to the new audio-only condition. An additional six children were tested but not included in the analysis, either due to failure to complete at least two practice trials (2 children), technical error (2) or distraction/refusal to participate during the test block (2).

2.2. Materials[2]

The experimental materials were obtained from the original de Carvalho, Dautriche & Christophe (2016) study. The authors created eight pairs of test sentences, with eight different pairs of noun-verb homophones. All possible meanings for each of the homophones were reported to be understood by children in the age range tested, according to the McArthur database for French (Kern, Langue, Zesiger & Bovet, 2010) and to a questionnaire addressed to parents (see e.g., de Carvalho et al., 2017). For each homophone, a pair of test sentences was created. One sentence used the homophone as a noun (noun prosody condition, e.g. [*La*DET *petite*ADJ ***ferme***NOUN][*lui plaît beaucoup*]), whereas the other used it as

[2] For a complete list of experimental materials, see our osf repository (https://osf.io/f9ecy/).

a verb (verb prosody condition, e.g. [*La*DET *petite*NOUN][***ferme***VERB *le coffre à jouets*]). Verb prosody sentences consistently presented a prosodic boundary before the homophone, whereas noun prosody sentences presented a boundary after this word. This was signaled by a significant lengthening of rhymes preceding each prosodic boundary (e.g., in /fɛʁm/, for the noun prosody, and in /pətit/, for the verb prosody) as well as a rising pitch contour towards the end of the prosodic units (see de Carvalho et al., 2016 for more details on the stimuli and acoustic analysis).

In addition to the experimental sentences, 11 filler sentences featuring target words that were unambiguously either a noun (e.g. [*Le*DET *bébé ADJ* ***oiseau***NOUN][*mange beaucoup*] "the baby **bird** eats a lot") or a verb (e.g. [*La*DET *maîtresse*NOUN][***parle***VERB *aux enfants*] "the teacher **talks** to the children") were also created. Each sentence was cut after the target word and 1000 ms of babble noise, created by superimposing the end of all filler sentences, was added. This babble noise was identical across all sentences. To create an analogous effect in the visual domain, in the audio-visual condition, at the offset of the target word, the video of the speaker lost contrast, became blurred, and trembled (making lip-reading fully impossible; see Figure 1, adapted from de Carvalho et al., 2016). The speaker's videos were presented inside a drawing of a television (because we told children that the television was broken, to explain why the videos became inaudible and distorted at the end). In the new audio-only condition, we used the same audio from the videos, but the visual information was replaced by a static image of a radio.

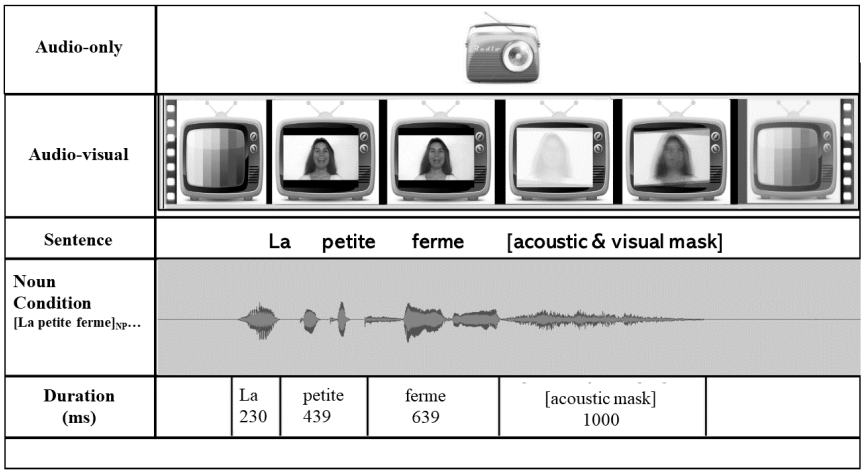

Figure 1. Example of a test sentence used in the experiment together with its waveform and the duration of each of the components, and what participants saw on the screen in each modality (audio-only vs. audio-visual). Figure adapted from de Carvalho, Dautriche & Christophe, 2016, p. 239.

Stimuli presentation was divided into two lists, so that each member of a given sentence pair would appear on a different list. Each list contained four sentences with noun prosody and four sentences with verb prosody, plus four filler sentences (two using familiar nouns and two with familiar verbs). Half of the participants were assigned to each list. The order of sentences within each list was pseudo-randomized, so that participants would see no more than two consecutive test sentences in a row and no more than two target words of the same category (noun or verb) in a row.

2.3. Procedure

Children were tested individually in a quiet room in their preschool. They sat in front of a computer and wore headphones to listen to the stimuli. At the beginning of the experiment, the experimenter would tell the child that they would listen to a woman on a television screen (audio-visual modality) or on a radio (audio-only modality), who would tell them stories. However, because the television/radio was broken, the child would not hear the end of the stories and would have to guess what the woman might have said. To elicit children's sentence completions, they were told that they were competing with other children on the screen, who were participating virtually, and that the child who completed more stories would win the game.

In each trial, an arrow appeared at the bottom right corner of the screen, surrounded by pictures of three children, which were presented as the participant's virtual competitors (see Figure 2).

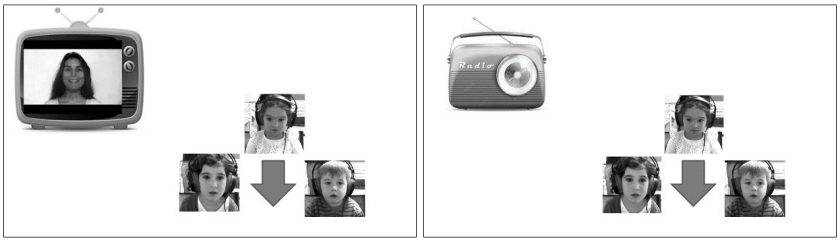

Figure 2. Example of the scenario used in Experiment 1. Children listened to the beginning of the ambiguous sentences either while watching the video of the lady speaking (audio-visual condition, scenario on the left) or while seeing a still picture of a radio (audio-only condition, scenario on the right). Finally, they repeated what they heard and completed the sentence.

At the beginning of each trial, the arrow would rotate to select one child to complete the sentence. If the arrow pointed down, it was the participant's turn to answer. The virtual competitors were chosen only to answer filler sentences, while the participant answered all test sentences. When the arrow pointed towards a virtual child on the screen, suggesting s/he was selected to respond, the

experimenter interacted with the virtual participant in the same way she interacted with the real participants (asking them to pay attention to the story "the lady will tell next" and then complete it). A pre-recorded sentence was played with the response given by the virtual participant. These sentences were previously recorded from children of the same age as the child participants. When the participant was selected, the experimenter asked them to pay attention to the video/audio that was coming up. Afterwards, they were asked to repeat what they heard and attempt to complete the sentence.

The experiment started with a practice block, in which children were presented exclusively with filler sentences. A maximum of eight filler sentences were randomly selected for this block. The virtual children were selected to complete the first two sentences, in order to introduce the participant to the task. The third and fourth trial chose the participant, and the subsequent trials alternated between the participant and the virtual children. The experimenter started the test phase as soon as the participant correctly completed two of the filler sentences in this practice block. If after eight practice trials the participant did not provide at least two correct answers, the experimenter ended the experiment and did not start the test trials.

The test phase was composed of eight test sentences and four filler sentences, half in the noun condition and half in the verb condition.

2.4. Data analysis

Participants' responses were recorded through the computer's microphone using *Audacity* (audacityteam.org), and were subsequently coded by the experimenter. The answers were coded as *noun answers* when the participant completed the sentence using the target homophone as a noun (e.g. "...is very nice"), or as *verb answers* when they used the target homophone as a verb in their completions (e.g. "... the window"). Since participants heard the sentences through headphones, the experimenter was blind to the condition of each test trial. For the child data, 79 out of the 384 responses were excluded from our analysis (48 for the audio-only condition), either because the child did not answer (n = 57), or because the answer was too ambiguous and consistent with both interpretations of the target word (n = 22). For instance, if a child completed the sentence "*la grande marche*" with "*face à ses yeux*" ("in front of their eyes"), this response was considered to be ambiguous because it could either mean "the big girl walks in front of their eyes" or "the big step in front of their eyes"[3]. For the adult data, two responses were excluded from analysis (both from participants in the audio-only condition) because the answer was ambiguous.

Since noun and verb responses were complementary, we chose the proportion of noun answers (1 = noun answer; 0 = verb answer) in each condition as our dependent measure. A generalized linear mixed-effects model with the proportion of noun answers as the dependent variable was performed. The presentation

[3] The prosody used by the child was not taken into account when coding the answers.

modality (audio-visual vs. audio-only) and the prosody conditions (noun prosody vs. verb prosody) were modeled as fixed effects, and participants and items as random intercepts[4]. Following the results from de Carvalho et al. (2016), we expected to find a significant effect of prosody, indicating that children and adults give more noun answers to sentences in the noun prosody condition than to sentences in the verb prosody condition. However, if the presence of facial coverbal speech gestures enhance children's perception of prosodic boundaries in the task, we would expect a significant interaction between prosodic condition and modality, showing that the effect of prosody is stronger for participants in the audio-visual modality than for participants in the audio-only modality.

2.5. Results

Figure 3 presents the average proportion of noun and verb answers for each prosody condition in each modality.

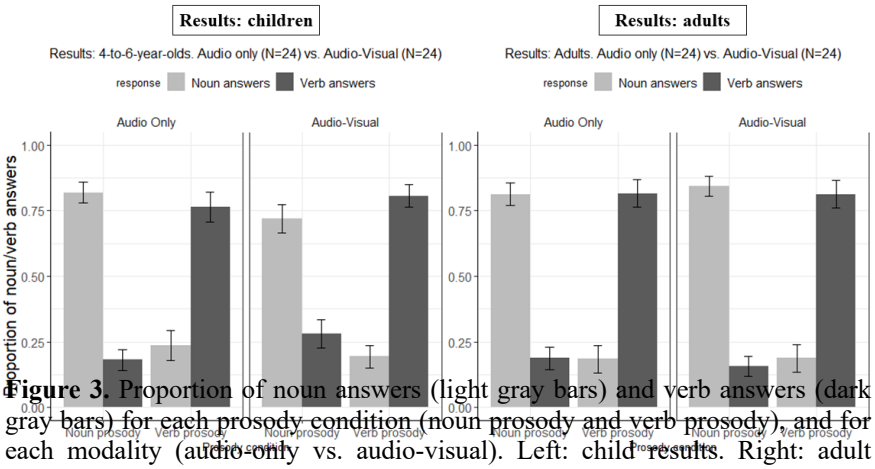

Figure 3. Proportion of noun answers (light gray bars) and verb answers (dark gray bars) for each prosody condition (noun prosody and verb prosody), and for each modality (audio-only vs. audio-visual). Left: child results. Right: adult results.

Children and adults gave overall more noun answers to sentences in the noun prosody condition than to sentences in the verb prosody condition (children: $\beta = -2.88$; SE = 0.47; $z = -6.15$; $p < 0.001$; adults: $\beta = -4.23$; SE = 0.49; $z = -8.67$; $p < 0.001$). However, no significant interaction between prosody condition and modality was found ($p > 0.8$ for both age groups).

[4] The maximal random effect structure that allowed the model to converge (Barr, Levy, Scheepers, & Tily, 2013).

2.6. Discussion

In the present study, we investigated whether facial coverbal speech gestures could impact children's ability to use prosody to constrain syntactic analysis. We replicated an oral completion task from de Carvalho, Dautriche & Christophe (2016) in two modalities: their original audio-visual modality, in which children had access to both auditory and visual cues to prosodic boundaries (through coverbal facial speech gestures of the speaker + prosodic cues), and a new audio-only modality, in which children listened to the recordings of the test sentences, but without simultaneous visual cues to prosodic boundaries. We believed that if children benefit from the presence of coverbal facial speech gestures to retrieve prosodic/syntactic boundaries in this task, participants in the audio-visual modality would perform better in interpreting the test sentences than the ones in the audio-only modality.

The overall results show that 4-to-6-year-old French-speaking children can successfully use prosodic boundary information to constrain their parsing in both audio-visual and audio-only modalities. Children gave significantly more noun completions to sentences in the noun prosody condition than to sentences in the verb prosody condition. This replicates de Carvalho et al.'s study, providing additional support for the hypothesis that young children can use prosodic boundary information to constrain syntactic analysis.

Regarding the comparison between modalities, the analysis did not return a significant interaction, showing that participants performed equally well regardless of whether they listened to the test sentences accompanied by the videos of the speaker or not. This suggests that the presence of coverbal facial speech gestures did not improve children's ability to use prosody to constrain parsing.

One possible objection to the observed similarity in performance between the audio-visual and audio-only modalities might stem from the lack of salience of the visual cues in the task. The speaker's videos occupied a limited portion of the screen, sharing space with unrelated visual elements, i.e., the pictures of the three virtual competitors surrounding the blue arrow (see Figure 2 above). Consequently, it is possible that the visual cues were not prominent enough for participants to effectively rely on them while processing the sentences. To explore this hypothesis, we designed a follow-up experiment (Experiment 2) where we expanded the size of the speaker's videos to cover the entire screen and eliminated the pictures of the virtual participants to avoid any competitor effect on participants' visual attention. If the absence of a distinction between modalities in Experiment 1 was indeed due to the size of the video and the presence of concurrent visual information, we anticipate observing a difference between the audio-visual and audio-only modalities in this revised version of the task.

3. Experiment 2

3.1. Participants

Child and adult data collection for Experiment 2 is still ongoing. So far, thirty-two adult participants, undergraduate students in psychology (M_{age} = 19.11 years old, range 17.2 to 24.10 years, 31 women) were included in the preliminary analysis of this task. Half of the participants were assigned to the audio-visual condition, and the other half was assigned to the audio-only condition. Preliminary results for child data are not included here, given the limited number of children tested up to the date of paper submission.

3.2. Materials

We used the same stimuli from Experiment 1. However, we removed the images of the virtual participants from the screen, and increased the size of the videos in the audio-visual condition so that they would occupy the whole screen. For the audio-only condition, we replaced the video with a picture of a radio also occupying the whole screen. The radio featured a visual display that illuminated and showed a loudspeaker icon during the playback of the sentence.

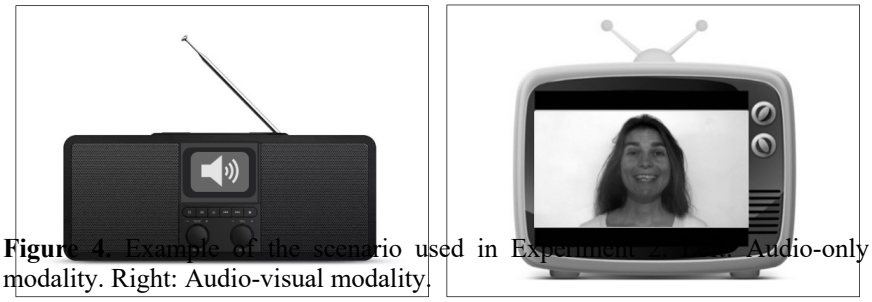

Figure 4. Example of the scenario used in Experiment 2. Left: Audio-only modality. Right: Audio-visual modality.

3.3. Procedure

The procedure closely mirrored that of Experiment 1, with the sole modification being the exclusion of the competition prompt from the task. Participants listened and completed all sentences across the experiment (a maximum of eight sentences from practice trials, eight test sentences and four filler sentences). As for Experiment 1, participants were directed to the test phase after successfully completing two sentences from the practice block.

3.4. Preliminary Results

Figure 5 presents the average proportion of noun and verb answers for each prosody condition and each modality. Adults gave overall more noun answers to sentences in the noun prosody condition than to sentences in the verb prosody condition (β = -5.27; SE = 1.13; z = -4.65; $p < 0.001$). However, again no significant interaction between prosody condition and modality was found ($p > 0.9$).

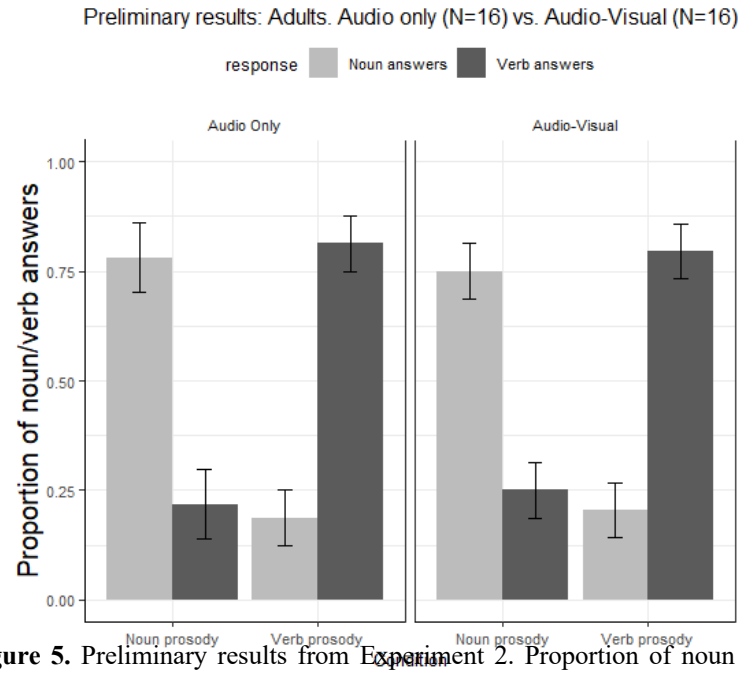

Figure 5. Preliminary results from Experiment 2. Proportion of noun answers (light gray bars) and verb answers (dark gray bars) for each prosody condition (noun prosody and verb prosody), and for each modality (audio-only vs. audio-visual).

4. General discussion

In the present study, we investigated the influence of natural coverbal facial speech gestures, such as head nods and eyebrow movements, on the perception of prosodic boundary information for sentence parsing. In an oral completion task, participants heard the beginnings of ambiguous sentences featuring noun-verb homophones, either with auditory prosodic cues accompanied by coverbal facial speech gestures (audio-visual condition) or auditory cues only (audio-only condition), and were then instructed to complete the sentences freely, offering insights into their interpretation of the homophones. The overall results showed that children and adults in both conditions were able to use prosodic boundary

information to constrain their syntactic analysis, as they assigned different syntactic categories to the ambiguous homophone depending on its position within the prosodic structure of the sentences heard. This aligns seamlessly with earlier research, reinforcing the notion that young children can employ prosodic boundary information to fine-tune syntactic analysis. Notably, our study replicates de Carvalho, Dautriche & Christophe's (2016) results across both audio-visual and audio-only modalities.

Regarding the comparison between modalities, the analysis of Experiment 1 yielded no significant difference in performance between participants in the audio-only and audio-visual conditions. This indicates that the presence of coverbal facial speech gestures did not enhance participants' ability to use prosody to constrain parsing in this particular task. Experiment 2 sought to determine whether amplifying the size of the videos on the screen and eliminating concurrent visual information would yield different results. Preliminary findings from adult participants indicated again no difference in performance between the two modalities, suggesting that the results observed in Experiment 1 might not be due to a lack of prominence in the speaker's videos in the audio-visual modality. However, the behavior of children in this version of the task remains unknown, as we did not yet complete data collection with this age group. Given the results of Exp 1, it is highly possible that children will again demonstrate adult-like performance in Experiment 2. Nevertheless, it remains to be seen whether the enlarged size of the speaker's videos may enhance children's attention to coverbal facial speech gestures in this task.

In summary, the obtained results suggest that coverbal speech gestures do not impact children and adults' ability to use prosodic information to constrain parsing of natural language. The removal of visual cues to prosodic boundaries did not decrease participants' ability to use prosodic information for sentence analysis in our task.

Future studies need to address a few lingering questions. Firstly, could coverbal speech gestures be more useful in contexts where acoustic cues to prosodic boundaries are less reliable or salient (e.g., in a noisy environment)? Can listeners benefit more from these additional cues to constrain parsing of more complex sentences? Secondly, given our focus on the natural production of coverbal speech gestures, we used videos in which the speaker was not directly instructed to produce such gestures, but only to talk as if she was naturally addressing a child. Therefore, it remains possible that if we created videos where the speaker is instructed to consistently integrate coverbal speech gestures with prosodic boundaries in a prominent manner, these gestures could be more useful as an additional cue to syntactic structure. Yet, considering the artificial nature of this condition, it would be hard to determine whether improved performance among participants exposed to coverbal speech gestures in this situation could be attributed to the inherent usefulness of these gestures in enhancing the perception of prosodic boundaries, or merely a result of participants attending to an unusual and thus attention-grabbing aspect correlating with phrase boundaries.

The present study adds to the growing body of evidence on the role of prosody in constraining parsing in young children, strengthening the notion that phrasal prosody is an important cue to syntactic structure during language acquisition. Beyond that, we contribute to studies on the role of coverbal speech gestures for sentence processing. Surprisingly, our study challenges previous results, by suggesting that the presence of natural coverbal facial speech gestures does not impact children and adults' ability to use prosody during parsing. Future studies will need to investigate more carefully the conditions under which coverbal facial speech gestures signaling prosodic boundaries might be useful for listeners during natural sentence processing.

References

Barr, Dale J., Levy, Roger, Scheepers, Christoph, & Tily, Harry J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Burnham, Denis & Dodd, Barbara (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 45*(4), 204-220.

de Carvalho, Alex, Dautriche, Isabelle, & Christophe, Anne (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental science, 19*(2), 235-250.

de Carvalho, Alex, Dautriche, Isabelle, Lin, Isabelle, & Christophe, Anne (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, *163*, 67-79.

de Carvalho, Alex, Kolberg, Leticia S., Trueswell, John, & Christophe, Anne. (2022, May). Cross-linguistic evidence for the role of phrasal prosody in syntactic and lexical acquisition. In *Speech Prosody 2022* (pp. 396-400).

de la Cruz-Pavía, Irene, Gervain, Judith, Vatikiotis-Bateson, Eric, & Werker, Janet F. (2020a). Coverbal speech gestures signal phrase boundaries: A production study of Japanese and English infant-and adult-directed speech. *Language Acquisition, 27*(2), 160-186.

de la Cruz-Pavía, Irene, Werker, Janet F., Vatikiotis-Bateson, Eric, & Gervain, Judith (2020b). Finding Phrases : The Interplay of Word Frequency, Phrasal Prosody and Co-speech Visual Information in Chunking Speech by Monolingual and Bilingual Adults. *Language and Speech, 63*(2), 264-291.

Esteve-Gibert, Núria & Prieto, Pilar (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language and Hearing Research, 56*(3), 850-864.

Gervain, Judith, Nespor, Marina, Mazuka, Reiko, Horie, Ryota, & Mehler, Jacques (2008). Bootstrapping word order in prelexical infants: A Japanese-Italian cross-linguistic study. *Cognitive Psychology, 57*, 56–74. doi: 10.1016/j.cogpsych.2007.12.001

Gervain, Judith, Sebastián-Gálles, Núria, Díaz, Begoña, Laka, Itziar, Mazuka, Reiko, Yamane, Naoto, Nespor, Marina, & Mehler, Jacques (2013). Word frequency cues word order in adults: Crosslinguistic evidence. *Frontiers in Psychology, 4*, 689. doi: 10.3389/fpsyg.2013.00689

Kern, S., Langue, J., Zesiger, P., & Bovet, F. (2010). Adaptations françaises des versions courtes des inventaires du développement communicatif de MacArthur-Bates. *Approche Neuropsychologique des Apprentissages chez l'Enfant, 107*(108), 217-228.

Kolberg, Leticia, de Carvalho, Alex, Babineau, Mireille, Havron, Naomi, Fiévet, Anne Caroline, Abaurre, Bernadete, & Christophe, Anne (2021). “The tiger is hitting! the duck too!” 3-year-olds can use prosodic information to constrain their interpretation of ellipsis. *Cognition, 213*, 104626.

Massicotte-Laforge, Sarah & Shi, Rushen (2015). The role of prosody in infants' early syntactic analysis and grammatical categorization. *Journal of Acoustical Society of America, 138*(October), 441–446. https://doi.org/10.1121/1.4934551

McGurk, Harry & MacDonald, John (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Morgan, James L. & Demuth, Katherine (Eds.). (2014). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Psychology Press.

Nespor, Marina & Vogel, Irene (1986). *Prosodic Phonology*. Dordrecht: Foris.

Rosenblum, Lawrence D., Schmuckler, Mark A., & Johnson, Jennifer A. (1997). The McGurk effect in infants. *Perception & psychophysics, 59*(3), 347-357.

Snedeker, Jessie & Yuan, Sylvia (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of memory and language, 58*(2), 574-608.