

A study of the McGurk effect in 4 and 5-year-old French Canadian children

Sophie Dupont

Jérôme Aubin

Lucie Ménard

University of Quebec in Montreal, Montreal, Canada

It has been shown that visual cues play a crucial role in the perception of vowels and consonants. Conflicting consonantal stimuli presented in the visual and auditory modalities can even result in the emergence of a third perceptual unit (McGurk effect). From a developmental point of view, several studies report that newborns can associate the image of a face uttering a given vowel to the auditory signal corresponding to this vowel; visual cues are thus used by the newborns. Despite the large number of studies carried out with adult speakers and newborns, very little work has been conducted with preschool-aged children. This contribution is aimed at describing the use of auditory and visual cues by 4 and 5-year-old French Canadian speakers, compared to adult speakers, in the identification of voiced consonants. Audiovisual recordings of a French Canadian speaker uttering the sequences [aba], [ada], [aga], [ava], [ibi], [idi], [igi], [ivi] have been carried out. The acoustic and visual signals have been extracted and analysed so that conflicting and non-conflicting stimuli, between the two modalities, were obtained. The resulting stimuli were presented as a perceptual test to eight 4 and 5-year-old French Canadian speakers and ten adults in three conditions: visual-only, auditory-only, and audiovisual. Results show that, even though the visual cues have a significant effect on the identification of the stimuli for adults and children, children are less sensitive to visual cues in the audiovisual condition. Such results shed light on the role of multimodal perception in the emergence and the refinement of the phonological system in children.

1. Introduction

Studies have shown that speech perception not only relies on the decoding of the waveform, but also on visual information transmitted by the speaker's jaw and lips positions. Interactions between speakers under unimodal communication conditions are frequent and most of the time successful. Telephonic conversations, where only the auditory modality is used, are good examples of such unimodal interactions. Considering speech in natural face-to-face interactions does not discern between the contribution of both modalities and their integration in the audiovisual speech perception; that's why, since the first data relative to this kind of experiment were published in 1976, the McGurk effect has constituted an efficient research paradigm for the study of audiovisual speech perception.

The McGurk effect was first introduced in an article by McGurk and MacDonald (1976) in which results of conflicting audiovisual stimuli percepts were presented. The first manifestation of this effect occurred when an auditory stimulus /ba/ was dubbed to a visual stimulus /ga/. Most of the time, people perceived /da/, a percept which resulted from the fusion of the auditory and visual information. Using the McGurk effect, many studies have since been conducted on adults but only a few investigated the emergence of this effect in speech development.

When McGurk and MacDonald (1976) discovered this audiovisual integration phenomenon, they decided to study its robustness in English speaking children. They tested 54 adults (18-40 years old), 21 pre-school aged children (3-4 years old) and 28 children of school age (7-8 years old) and presented them stimuli in two conditions : audiovisual (for example, A (audio) /ba/ - V (visual) /ga/, A /ga/ - V /ba/, A /pa/ - V /ka/ and A /ka/ - V /pa/) and visual-only. The authors found that adult subjects were more influenced by the visual input than were the two other groups. The description of this phenomenon by the authors also showed that percepts reflecting the absence of any audiovisual integration were audio percepts for child subjects, and visual percepts for adult subjects.

In a 1978 study, MacDonald and McGurk sought to explain the phenomenon of the conflicting auditory and visual information by testing the manner-place hypothesis, according to which, in a face-to-face situation, manner of articulation in consonants is best recognized by ear, while articulation place is best integrated by eye. In their experiment, 44 English speaking adults participated in a perception test where they had to repeat the perceived utterance produced by a female speaker presented on a screen. Stimuli were conflicting

dubbings of the consonants /p, b, t, d, k, g, m, n/ in an /a/ vocalic context. The manner-place hypothesis was validated with percepts for which the audio part of the stimuli consisted of labial consonants and the visual part consisted of non-labial stimuli (for example, A /ma/ - V /ga/ stimuli were mostly identified as /na/), but was “less satisfactory with respect to nonlabial sound/labial lips combinations” (MacDonald et McGurk 1978: 256) such as A /ga/ - V /pa/ stimuli, which were identified as /ga/. This study showed the general effect of vision in speech perception in face-to-face interactions.

In 1984, Massaro conducted a study concerning the evaluation of the integration of information in speech perception, which aimed at studying the developmental aspects of speech perception. He compared results from 11 children to those from 11 adults in a phonemic identification task, and found that the children showed half the visual influence of the adults. He mentioned that children seemed sensitive to the correspondence of visual and auditory information, but auditory information had a greater influence on their categorical perception in language acquisition.

Rosenblum et al (1997) investigated the McGurk effect in 5-month-old babies. They wanted to show that speech representation of babies was amodal and that babies processed the audiovisual input by making a supramodal association with the audiovisual input instead of an association with the phonemic identity. By creating three types of audiovisual stimuli (conflicting, non-conflicting and audio-only) and by using an habituation technique, they measured fixation time and found a significant difference between the fixation time of the stimulus A /da/ - V /va/ and the stimulus A /va/ - V /va/, but not between A /ba/ - V /va/ and the stimulus A /va/ - V /va/. Since they ensured that a general preference of the babies for the A /da/ - V /va/ stimuli or an auditory similarity between the stimuli could not account for this significant difference, they concluded that they had “observed evidence for a McGurk-type effect in 5-month-old infants”.

More recently, Robert-Ribes et al. (1998) identified the complementarity and synergy between vision and hearing as two factors that influence the effectiveness of audiovisual speech perception. Complementarity is related to the manner-place hypothesis of MacDonald and McGurk (1978), in that it explains that the manner of articulation is best transmitted by the auditory channel and that place of articulation is best transmitted by the visual channel. Synergy is a property related to the perception enhancement caused by the interaction between both the auditory and the visual modalities; audiovisual perception is thus always better than visual-only perception and auditory-only perception. The authors also used the notion of viseme, first introduced by

Fisher (1968) to refer to “visual phonemes” and this notion will be relevant to the present study.

Altogether, these studies suggest that audiovisual perception development is not equally weighted during speech development: babies and young children would rely less than adults on visual cues. But since we know that they can use them in specific perception tasks, like matching acoustical signals to visual patterns as early as 4 and 5-months old (Kuhl and Metzoff, 1984) and that attention can be monitored during the perception tests (Massaro, 1984; Tiipana et al., 2004), we could hypothesize that children use as many cues as available during speech development, but to an extent specific to each child. The goal of this research is thus to describe the role of auditory and visual cues during speech development in French Canadian speaking subjects.

2. Method

2.1. Stimuli Recordings

Audiovisual recordings of a French Canadian adult speaker producing symmetric VCV sequences were carried out, using a *Panasonic* Mini-DV camcorder. Vowels /a/ and /i/ and consonants /b/, /d/, /g/, and /v/ were used to construct the following sequences, repeated three times: [aba], [ada], [aga], [ava], [ibi], [idi], [igi], and [ivi]. Speech rate, intonation, and intensity were controlled by the experimenters in order to maintain them at constant levels. As can be observed in Figure 1, a close-up on the lower part of the speaker’s face was made. After the second vowel of each sequence, the speaker was told to return to a neutral position, lips closed.

Those sequences were digitized at a frame rate of 29.94 frames per second and at a sampling frequency of 44.1 KHz. Audiovisual data were imported using *Imovie* (Apple) and bisyllabic sequences were edited. From the three repetitions recorded for each sequence, we kept only the one with the highest acoustic quality and movement clarity; each sequence lasted about 2800 ms.



Figure 1: Speaker's image after producing a sequence.

2.2. *Audio and Video Processing*

Recorded stimuli were edited in order to create the four following conditions. Bimodal non-conflicting stimuli were composed of the original image and sound of the speaker, repeating the 8 VCV sequences ([aba], [ada], [aga], [ava], [ibi], [idi], [igi], [ivi]) and did not require any editing. The same stimuli were used to construct the 8 visual-only stimuli, in which the sound was turned off. Unimodal audio condition stimuli consisted of the audio track of the original audiovisual non-conflicting stimuli.

Bimodal conflicting stimuli were constructed to generate consonantal conflicts between the acoustic and visual signals. *Adobe Premiere Pro 7.0* was used to separate the signals and superimpose them. Prior to this dubbing, closing and opening times of the mouth and lips were measured to ensure optimal synchronization between the audio and the video tracks. The combination of each of the four consonants in the auditory modality to every different consonant in the visual modality into the two vocalic contexts gave rise to a total of 24 stimuli.

2.3. *Perception Test*

2.3.1. *Subjects*

Eight children (all female) from four years and three months to five years and nine months (mean of four years and seven months) participated in this study. They were recruited from a daycare center in Montreal. 10 women, from 22 to 31 years old (mean of 25 years old) also served as subjects. All adult participants were enrolled in a Linguistics degree at the University of Quebec in Montreal. Children received a certificate of participation as a gift for their

contribution and adults received no compensation. All subjects were native French Canadian speakers and presented no vision or hearing problems. Most of the subjects did not know the goal of the experiment and did not seem to be conscious of the conflicting nature of the presented stimuli. Three adult subjects noticed the conflicting nature of the stimuli and believed this was done in order to disturb their perception. Percepts of those subjects were kept since McGurk and MacDonald (1976) mentioned the fact that being aware of the way stimuli were constructed did not inhibit the manifestation of the McGurk effect. Furthermore, informal comparisons showed no difference between the results of the two groups (those who were aware of the conflict and those who were not).

For both groups of subjects (children and adults), conflicting and non-conflicting stimuli were grouped in the bimodal condition category. Unimodal visual and auditory conditions, as well as bimodal conditions, were presented in three separated sessions to participants. Unimodal conditions each consisted of 8 stimuli, while the bimodal condition had 32 stimuli. Stimuli were randomized and presented once, separated by a four-second silent black screen.

2.3.1.1. *Adults*

Using *Windows Media player 9.0*, stimuli were presented to adult subjects in a perception test where the task was to be attentive to the movie presented on the screen and to the sound presented in the earphones and to write down the perceived utterance. In addition to the four-second black screen between each stimulus, subjects could also press pause between sequences to allow them more time to write down their responses. In order to avoid a lack of attention due to the task of writing and watching the screen at the same time, we asked the last three subjects who participated in the test to repeat the perceived utterance, instead of writing it down. A random list of stimuli was presented in the following order: audiovisual (conflicting and non-conflicting mixed) stimuli, audio stimuli, and visual stimuli. This test took place in a quiet room and lasted about fifteen minutes.

2.3.1.2. *Children*

For children, this test was presented as a game in which they were invited to be attentive to what was presented on the screen and to what they were hearing in the speakers. Their task was to repeat what they thought the speaker had just said. Two experimenters were present to write down the children's percepts, providing an inter-judge agreement and ensuring that the children were paying attention to what they were being presented. Experimenters monitored the test

by pressing the pause button between each utterance. A random list of stimuli was presented in the following order: audio stimuli, visual stimuli and audiovisual stimuli.

2.4. Data processing

For each condition, classification criteria were established for the consonant parts of the percepts; vocalic parts were not analyzed. Results are based on the mean numbers of percepts for a given group at a given condition for a given classification criterion. Then, results obtained from the children's data to a given classification criterion were compared to those of the adults for the same criterion. One-way ANOVAs were computed with Statistica 6.1. Student's t-tests were used to measure the perception difference of non-conflicting stimuli among a same group of subjects, but for different conditions (audio-only, visual-only and audiovisual). Student's t-tests were also performed to measure the perception difference of conflicting stimuli among a given type of percept (audio, visual, fusion, combination, other) between the two groups of subjects. The difference observed between the children and the adults will be described in these statistical terms, but remarks on specific percepts will be allowed only if at least 2 responses are similar. Indeed, since 16 answers were collected with the child subjects to each stimulus (8 subjects * 2 vocalic contexts) and 20 answers with the adults subjects (10 subjects * 2 vocalic contexts), a percept has to have been proposed twice by a subject or once by two subjects to be considered; percentage of 13% for children (2 percepts/16 responses) and 10% for adults (2 percepts/20 responses) is thus required.

3. Results

The principal goal of the following analysis was to describe the use of auditory and visual cues in the perception of four French Canadian consonants (/b/, /d/, /g/, /v/) by preschool-aged children and adults. To do so, we examined the results of the auditory-only and visual-only conditions to assess the use of each modality by the listeners. The analysis of the results of the conflicting stimuli revealed an influence of the visual component of the stimuli for both preschool-aged child and adult subjects, but the influence in children was clearly lower.

3.1. Control conditions

The analysis of the responses given to the unimodal (audio-only and visual-only conditions) and non-conflicting bimodal stimuli aimed at measuring: 1) the listeners' ability to identify the sequences in audio only and visual only

conditions and 2) the extent to which visual cues presented together with audio information improves intelligibility. Results can also provide an assessment of the quality of the audio components of the consonants as good exemplars of their phonetic categories.

3.1.1. Correct responses

For the auditory-only and the non-conflicting auditory-visual stimuli, percepts for which place and manner of articulation were the same as the presented stimuli have been treated as correct responses. For the visual-only condition, the responses for which the percepts and the stimuli shared the same place of articulation have been considered as correct responses. Table 1 presents examples of correct responses given by the two groups of subjects for each condition. Percepts which did not meet the criteria mentioned above have been classified as “others”.

Table 1: Some examples of the classification of the percepts in the control conditions.

Control conditions	Stimuli		Percepts	
	Audio component	Visual component	Correct responses	Others
Auditory-only	/b/	-	/b/	/d/
Visual-only	-	/d/	/d/, /t/	/b/
Auditory-visual	/g/	/g/	/g/	/v/

Figure 2 shows the average values of correct responses given by the adult subjects (top graph) and child subjects (bottom graph) for each consonant in the three control conditions. It is noteworthy that a ceiling effect is present for the /b/, /d/, and /g/ adults’ results, which all reach a proportion of correct responses of at least 85.0%. The slightly lower performances observed for the /v/ stimuli in the audiovisual and audio-only conditions cannot be explained by specific acoustic characteristics since no effect was observed on the identification of the conflicting stimuli having /v/ as audio part. Results of ANOVA did not reveal any significant difference in adults between any of the three control conditions. It shows that the audio and the visual components of the stimuli were intelligible and that the adults were competent in the tasks of listening and lip-reading.

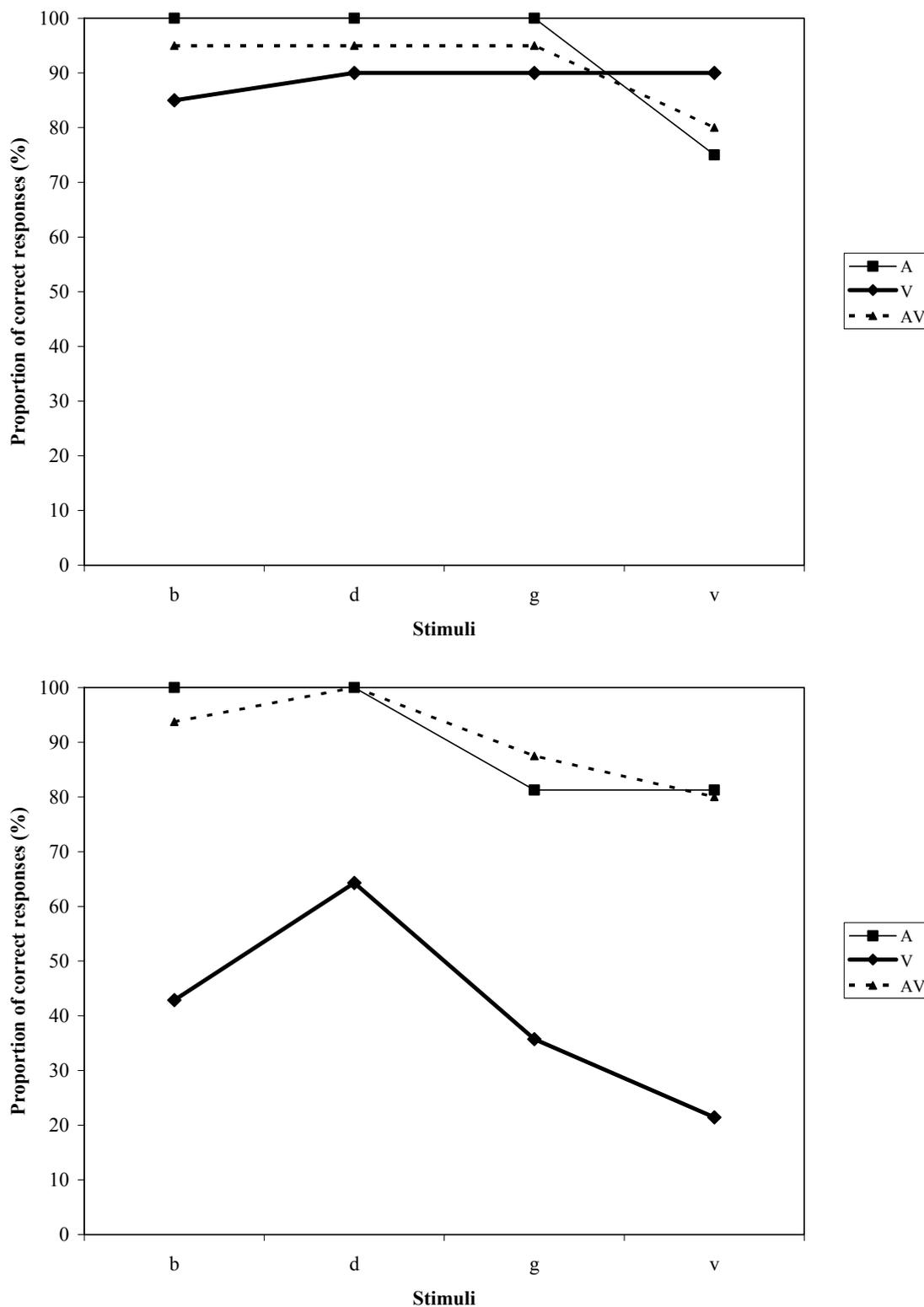


Figure 2: Proportion of correct answers for each consonant given by the adult subjects (top) and children subjects (bottom) in the three control conditions: visual-only (thick black line), auditory-only (thin black line) and non-conflicting auditory-visual conditions (dotted line).

In the bottom graph, we can observe the same kind of ceiling effect in the children results for the auditory-only and the auditory-visual conditions. However, a t-test revealed significant differences between the visual-only and the auditory-only conditions ($t=4.97$ $p<0.005$) and between the visual-only and the non-conflicting auditory-visual conditions ($t=4.75$ $p<0.01$). The performance of children in lip-reading was clearly poorer than the adults' and than their own performance in the two other control conditions. This visual-only condition was undoubtedly the most difficult one for children of this age since this task is rather unusual and infrequent in their day-to-day life. Since they were not given any choice of answer, we could not compare their results with a specific chance level; on the other hand, they gave correct answers between 21.4% and 64.2% of the time so we considered that they were sensitive to a certain extent to the speaker's articulatory movements.

3.2. *Conflicting conditions*

The analysis of the responses given to the conflicting auditory-visual stimuli aimed at studying the influence of the auditory and the visual modalities in multimodal speech perception. The identification of the conflicting stimuli gave rise to five classes of percepts: audio, visual, combination, fusion and other percepts. Data will be described and compared across the groups of child and adult subjects. Table 2 presents examples of each of the five classes of percepts given by the two groups of subjects to the conflicting stimuli.

Table 2: Some examples of the 5 classes of the percepts of the conflicting stimuli.

Stimuli		Percepts					
Audio component	Visual component	Audio	Visual	Combinations		Fusions	Others
				Strict	Extended		
/b/	/d/	/b/	/d/	-	-	-	/t/, /vg/
/b/	/g/	/b/	/g/	-	-	/d/, /θ/	-
/g/	/v/	/g/	-	/vg/, /gb/	/bg/	/d/	/f/
/v/	/d/	/v/	/d/	-	/bv/, /zd/	/θ/	/t/

3.2.1. *Audio percepts*

Percepts for which place and manner of articulation were similar to those of the audio component of the conflicting stimulus have been considered as audio percepts. They emphasize the absence of conflicting perception between the auditory and visual modalities. Figure 3 illustrates the means of the proportion of conflicting stimuli identified as audio percepts.

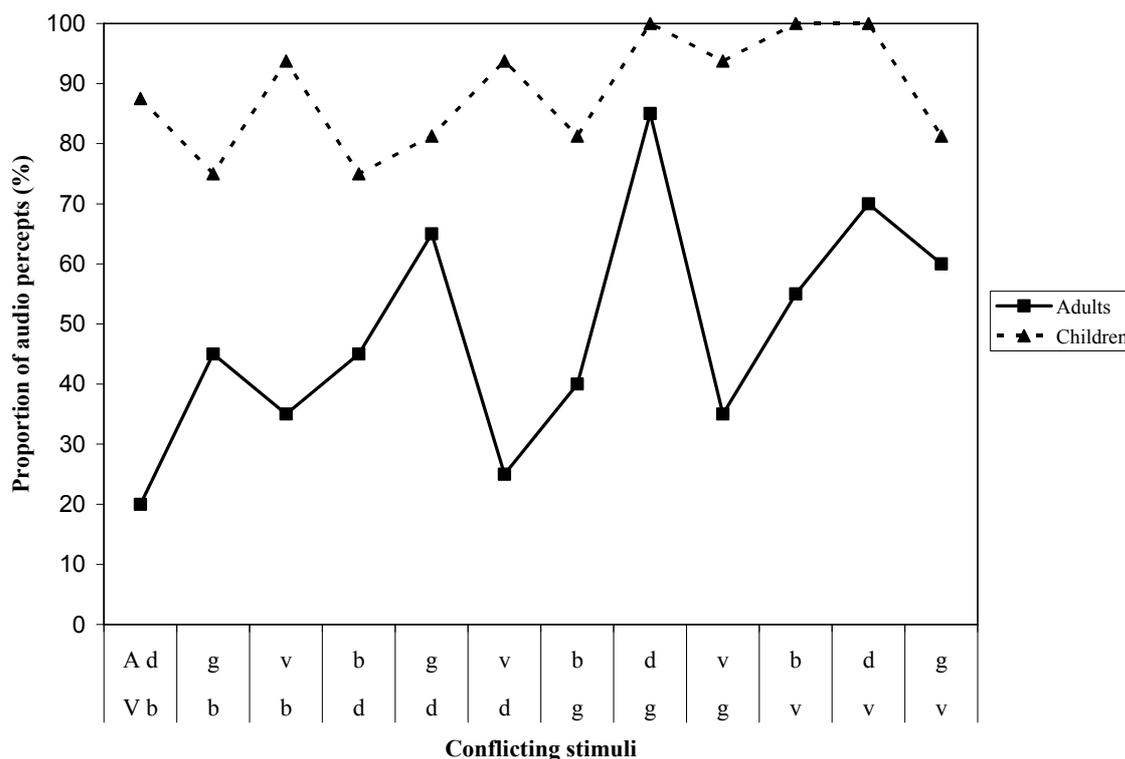


Figure 3: Proportion of audio percepts given by the adult subjects (black line) and child subjects (dotted line) for the 12 conflicting stimuli.

The proportion of audio percepts is significantly higher for children than for adults, as revealed by a one-way ANOVA ($F(1,22)=42.10, p<0.001$). These results are related to those found in the control conditions in children, since they already seemed to be more sensitive to auditory cues. We can also observe a tendency in the children's results: stimuli for which places of articulation of the auditory and visual components are the closest seem to give rise to more audio percepts. This tendency shows that children are more sensitive to auditory cues when less conflicting visual information is present. Even if the adults show a smaller proportion of audio percepts than children, the same tendency (for example for stimuli like A/g/-V/d/) is present.

3.2.2. Visual percepts

Visual percepts are those for which place and manner of articulation are similar to the visual component of the conflicting stimuli. Such percepts demonstrate a major dominance of the visual modality over the auditory modality. Figure 4 illustrates the means of the identification of the conflicting stimuli as visual percepts.

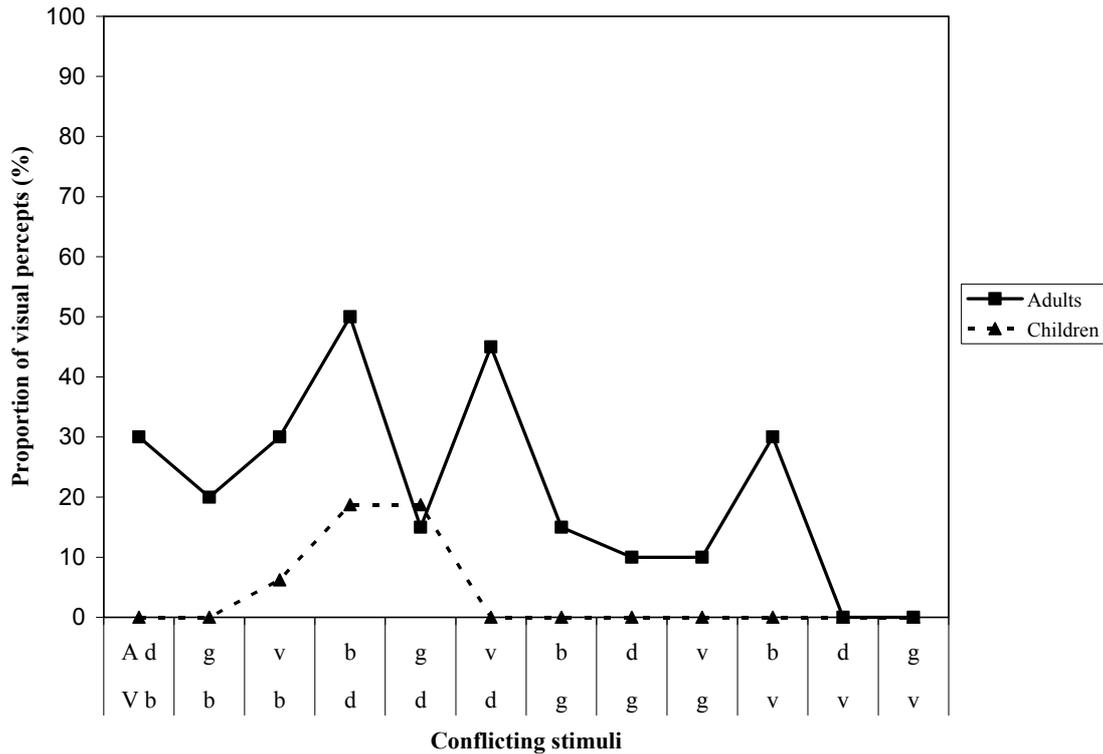


Figure 4: Proportion of visual percepts given by the adult subjects (black line) and children subjects (dotted line) for the twelve conflicting stimuli.

The proportion of visual percepts is significantly lower for children than for adults ($F(1,22)=11.90$, $p<0.01$). Since the results in the control conditions and the results of the audio percepts revealed the dominance of the auditory modality of children in speech perception, it is not surprising to observe such an effect. According to the criterion defined in section 2.4, children showed visual percepts in two contexts (A /b/ - V /d/ and A /g/ - V /d/), both of them including the dental /d/ visual consonant, which is surprising since dental place of articulation is not the most visible one. Note that the visual percept associated with A /v/ - V /b/ corresponds to only one response and could hardly be significant.

Visual percepts emphasize the important influence of the visual modality in speech perception; stimuli were presented in a quiet room and no surrounding noise could have distracted the subjects in their auditory perception. Thus, these visual percepts emphasize the dominance of the use of the visible articulatory gestures (visual cues) over the acoustical characteristics (auditory cues) of the conflicting stimuli.

3.2.3. Combination percepts

As can be seen in Table 2, combination percepts have been divided into two subclasses: strict combinations and extended combinations. Strict combination percepts include both the auditory and the visual consonants of the conflicting stimuli. Extended combination percepts are composed of two phonemes: one directly taken from the auditory or the visual component, and another one which is not directly taken from the components. Figure 5 shows the means of the proportion of conflicting stimuli perceived as combinations (strict and extended together).

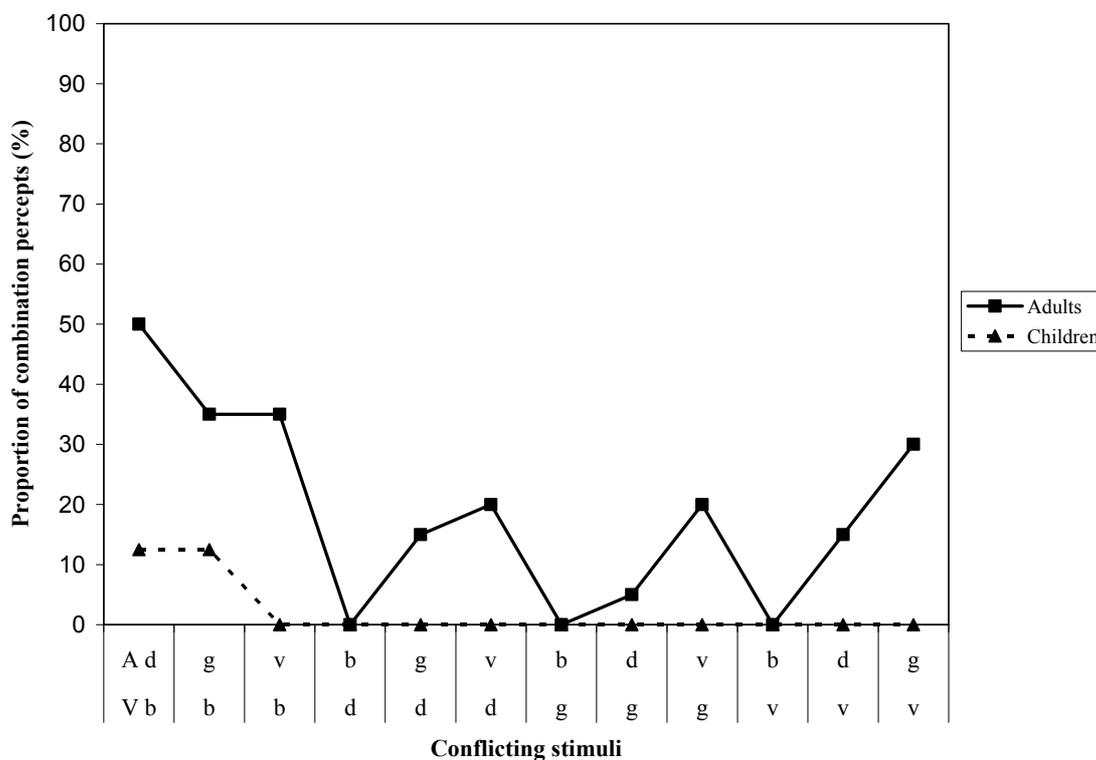


Figure 5: Proportion of combination percepts given by the adult subjects (black line) and child subjects (dotted line) for the twelve conflicting stimuli.

The proportion of combination percepts is significantly lower for children than for adults ($F(1,22)=11.58, p<0.01$). Since children showed only strict combinations, we compared the number of strict combination percepts of the children to the number of both strict and extended combinations percepts of the adults. Children showed combination percepts in two contexts: A /d/ - /V/ b/ and /A/ g/ - V /b/ and these contexts are those in which adults showed the highest proportion of combination percepts as well. This result suggests that the articulatory movements of the bilabial /b/ gave rise to more combination percepts when combined with a less visible visual consonant (dental /d/ and

velar /g/). On the other hand, some of the contexts in which adults gave the highest proportion of combination percepts include the audio velar /g/ consonant (A /g/ - V /b/ and A /g/ - V /v/). This result is not in agreement with McGurk and MacDonald (1976) nor MacDonald and McGurk (1978) who found that combination percepts were more frequent when dental and bilabial consonants were presented in the audio modality.

Combination percepts demonstrate the simultaneous influences of both the information provided by the auditory and the visual modality so it is interesting to note that children seem to be sensitive to the visual information, though to a smaller extent than the adults.

3.2.4. Fusion percepts

Fusion percepts are the most typical percepts of the McGurk Effect. When the subjects identified the conflicting stimuli by using only one phoneme and that phoneme was not that of the auditory part nor of the visual part, but instead, an “integration” percept, the responses were classified as fusions. Figure 6 shows the means of the identification of the conflicting stimuli as fusions.

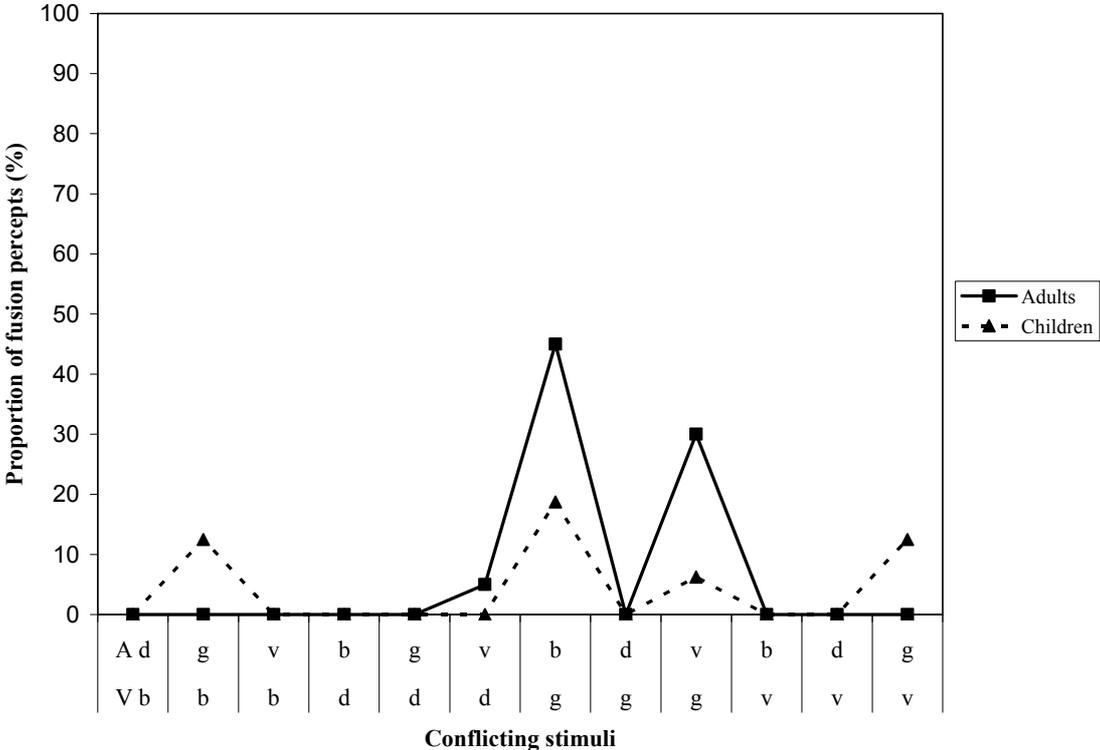


Figure 6: Proportion of fusion percepts given by the adult subjects (black line) and child subjects (dotted line) for the 12 conflicting stimuli.

The proportion of fusion percepts of children is not significantly different from that of adults. However, it is interesting to note that children showed fusions in three contexts (A /d/ - V /b/, A /b/ - V /g/ and A /g/ - V /v/) and adults in only two contexts (A /b/ - V /g/ and A /v/ - V /g/). The A /b/ - V /g/ stimuli are the ones which gave rise to the most /d/ or /θ/ fusion percepts. These percepts show that both children and adults were sensitive to visual cues and that they integrated them to the auditory cues giving rise to a fused perceived consonant.

3.2.5. Other percepts

All percepts given by the subjects which did not meet the conditions described previously have been classified as other percepts. Figure 7 shows the means of the identification of the conflicting stimuli as other percepts.

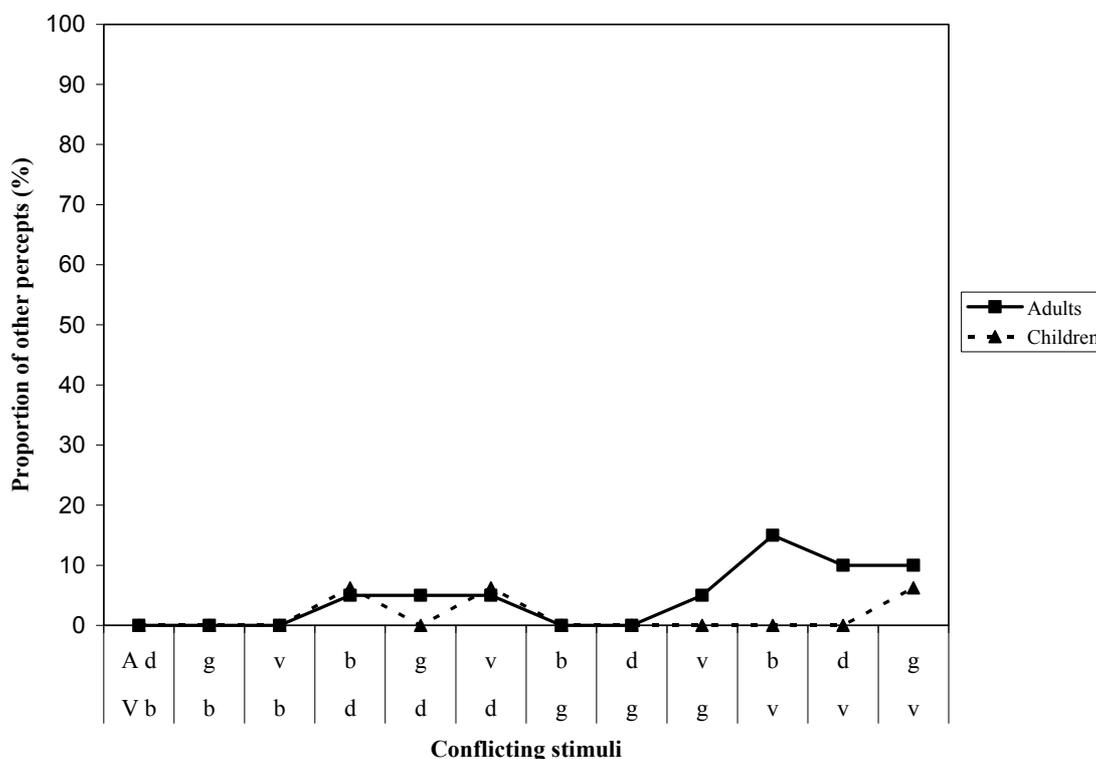


Figure 7: Proportion of other percepts given by the adult subjects (black line) and child subjects (dotted line) for the twelve conflicting stimuli.

The proportion of other percepts is not significantly different from that of adults. Some of the other percepts cannot be recovered by the auditory or the visual cues (ex: A/b/ - V/d/ identified as /vg/) and could be considered as random responses or errors. However, the interpretation of some other percepts can be related to the visual component. For example, 36.4% of the other percepts given

by the adults consisted in the devoiced visual component of the conflicting stimuli.

It is also interesting to note that the children did not give more ‘other’ percepts than adults; we could have hypothesized that they would have given more random responses due to a lack of attention or to the difficulty of the task, but it seems not to have happened.

4. Discussion and conclusion

Many differences between the child subjects’ percepts and the adult subjects’ percepts appeared to be significant. Overall, the children subjects were more sensitive to auditory cues. Their low performance in lip-reading and their greater number of audio percepts in response to the conflicting stimuli are closely akin to Massaro’s postulate (1984) on the greater influence of the auditory information on the perception of the phonetic categories in language acquisition.

Even if the children subjects have been much more influenced by the auditory information, they have nonetheless showed the McGurk effect in some contexts, but the effect was lower than in adults and much more variable. These results have to be taken into account in a description of the role of the vision and audition in the phonological development in children.

But many questions still need to be investigated to provide a coherent description. The magnitude of the McGurk effect across development seems to be non-linear; Rosenblum’s (1997) data collected by a head-turn procedure with 5-month-old infants suggest that they are sensitive to the McGurk effect and data collected in preschool-aged children like ours (McGurk and MacDonald, 1976; Massaro 1984) suggest it too but to a smaller extent. Finally, evidence has widely demonstrated the McGurk effect in adults. We could have hypothesized that since infants are sensitive to visual cues and since children are in the process of acquiring their phonological categories, they would use as many cues as available. However, this hypothesis is not supported by the present data.

Further work should be conducted with more subjects over a wider range of age to try to localize some kind of “McGurk effect critical period” and to study what would underlie such a period. Finally, the relationship between speech perception and production could be explored in order to compare the use of auditory cues and visual articulatory movements in children and adults. Bilingualism might be a way to explore this issue. It may be of interest to

attempt to quantify second language fluency in some way, to determine whether some correlation might exist with observed McGurk percepts.

Acknowledgment

We would like to thank our speaker, Raoul Bugueno, for his patience. This work greatly benefited from the discussions and suggestions of the participants at the French-German summerschool on "Cognitive and physical models of speech production, perception and perception-production interaction", and from the thorough review by Pascal Perrier, Mark Tiede and Susanne Fuchs. This work was supported by FQRSC and CRLMB grants.

References

- Fisher, C. L. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, 11: 796-804.
- Kuhl, P. K. & Meltzoff, A. N. (1984). The intermodal representation of speech in infants, *Infant Behavioral Development*, 7: 361-381.
- MacDonald, J. M. & McGurk, H. (1978). Visual influences on speech perception process. *Perception and Psychophysics*, 24: 253-257.
- Massaro, D. W. (1984). Children's Perception of Visual and Auditory Speech. *Child Development*, 55: 1777-1788.
- McGurk H. & MacDonald, J. M. (1976). Hearing lips and seeing voices. *Nature*, 264: 746-748.
- Robert-Ribes, J., Schwartz J-L, Lallouache T. & Escudier, P. (1998). Complementary and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise. *Journal of Acoustical Society of America*, 103 (6): 3677-3689.
- Rosenblum, L. D., Schmuckler M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, 59(3): 347-357.
- Tiipana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, 16: 457-472.