# Auditory and Somatosensory Interaction in Speech Perception in Children and Adults

*Paméla Trudeau-Fisette[1,2]\*, Takayuki Ito[3,4] and Lucie Ménard[1,2]*

[1]*Laboratoire de Phonétique, Université du Québec à Montréal, Montreal, QC, Canada, [2]Centre for Research on Brain, Language and Music, Montreal, QC, Canada, [3]GIPSA-Lab, CNRS, Grenoble INP, Université Grenoble Alpes, Grenoble, France, [4]Haskins Laboratories, Yale University, New Haven, CT, United States*

Multisensory integration (MSI) allows us to link sensory cues from multiple sources and plays a crucial role in speech development. However, it is not clear whether humans have an innate ability or whether repeated sensory input while the brain is maturing leads to efficient integration of sensory information in speech. We investigated the integration of auditory and somatosensory information in speech processing in a bimodal perceptual task in 15 young adults (age 19–30) and 14 children (age 5–6). The participants were asked to identify if the perceived target was the sound /e/ or /ø/. Half of the stimuli were presented under a unimodal condition with only auditory input. The other stimuli were presented under a bimodal condition with both auditory input and somatosensory input consisting of facial skin stretches provided by a robotic device, which mimics the articulation of the vowel /e/. The results indicate that the effect of somatosensory information on sound categorization was larger in adults than in children. This suggests that integration of auditory and somatosensory information evolves throughout the course of development.

**Keywords: multisensory integration, speech perception, auditory and somatosensory feedback, adults, children, categorization, maturation**

## INTRODUCTION

From our first day of life, we are confronted with multiple sensory inputs such as tastes, smells, and touches. Unconsciously, related inputs are combined into a single input with rich information. Multisensory integration (MSI), also called multimodal integration, is the ability of the brain to assimilate cues from multiple sensory modalities that allows us to benefit from the information from each sense to reduce perceptual ambiguity and ultimately reinforce our perception of the world (Stein and Meredith, 1993; Stein et al., 1996; Robert-Ribes et al., 1998; Molholm et al., 2002). MSI holds a prominent place in the way that information is processed, by shaping how inputs are perceived. This merging of various sensory inputs into common neurons was typically assumed to occur late in the perceptual process stream (Massaro, 1999), but recent studies in neurophysiology have even demonstrated that MSI can occur in the early stages of cortical processing, even in brain regions typically associated with lower-level processing of uni-sensory inputs (Macaluso et al., 2000; Foxe et al., 2002; Molholm et al., 2002; Mishra et al., 2007; Raij et al., 2010; Mercier et al., 2013).

While some researchers have suggested that an infant's brain is likely equipped with multisensorial functionality at birth (Bower et al., 1970; Streri and Gentaz, 2004), others have suggested that MSI likely develops over time as a result of experiences (Birch and Lefford, 1963;

Yu et al., 2010; Burr and Gori, 2011). Several studies support the latter hypothesis. For example, studies have demonstrated that distinct sensory systems develop at different rates and in different ways, which suggests that several mechanisms are implicated in MSI depending on the type of interactions (Walker-Andrews, 1994; Gori et al., 2008; Burr and Gori, 2011; Dionne-Dostie et al., 2015). For example, researchers have reported that eye-hand coordination, a form of somatovisual interaction, can be observed in infants as young as a week old (Bower et al., 1970), and audiovisual association of phonetic information emerges around 2 months of age (Kuhl and Meltzoff, 1982; Patterson and Werker, 2003), but audiovisual integration in spatial localization behavior does not appear before 8 months of age (Neil et al., 2006).

Ultimately, although it is still unclear whether an innate system enables MSI in humans, data from infants, children, and adults suggest that unimodal and multimodal sensory experiences and brain maturation enables the establishment of efficient integration processing (Rentschler et al., 2004; Krakauer et al., 2006; Neil et al., 2006; Gori et al., 2008; Nardini et al., 2008; Hillock et al., 2011; Stein et al., 2014) and that multisensory tasks in school-aged and younger children are executed through unimodal dominance rather than integration abilities (McGurk and Power, 1980; Hatwell, 1987; Misceo et al., 1999; Burr and Gori, 2011). Moreover, according to the intersensory redundancy hypothesis, perception of multimodal information is only facilitated when information from various sources is redundant, and not when the information is conflicting (Bahrick and Lickliter, 2000, 2012).

Multimodal integration is crucial for speech development. According to the associative view, during infancy, the acoustic features of produced and perceived speech are associated with felt and seen articulatory movements required for their production (Kuhl and Meltzoff, 1982; Patterson and Werker, 2003; Pons et al., 2009; Yeung and Werker, 2013). Once acoustic information and proprioceptive feedback information are strongly linked together, this becomes part of an internal multimodal speech model (Guenther and Perkell, 2004; Tourville and Guenther, 2011; Guenther and Vladusich, 2012).

MSI can sometimes be overlooked in speech perception since speakers frequently have one dominant sensory modality (Hecht and Reiner, 2009; Lametti et al., 2012). However, even though audition is the dominant type of sensory information in speech perception, many researchers have suggested that other sensory modalities also play a role in speech processing (Perrier, 1995; Tremblay et al., 2003; Skipper et al., 2007; Ito et al., 2009; Lametti et al., 2012). The McGurk effect, a classic perceptual illusion resulting from incongruent simultaneous auditory and visual cues about consonants clearly demonstrates that information from multiple sensory channels is unconsciously integrated during speech processing (McGurk and MacDonald, 1976).

In the current study, we examined the integration of auditory and somatosensory interaction in speech perception. Previous research has suggested that to better understand how different types of sensory feedback interact in speech perception, we need to better understand how and when this becomes mature.

Hearing is one of the first sensory modalities to emerge in humans. While still *in utero*, babies can differentiate speech from non-speech and distinguish variability in speech length and intensity (for a review on auditory perception in the fetus, see Lecanuet et al., 1995). After birth, babies are very soon responsive to various rhythmic and intonation sounds (Demany et al., 1977) and can distinguish phonemic features such as voicing, manner, and place of articulation (Eimas et al., 1971). Specific perceptual aspects of one's first language, such as sensitivity to phonemes and phonotactic properties, are refined by the first year of life (Kuhl, 1991). Although auditory abilities become well established in the early years of life, anatomical changes and experiences will guide the development of auditory skills throughout childhood (Arabin, 2002; Turgeon, 2011).

Little is known about the development of oral somatosensory abilities in typically developing children. Yet, some authors have worked on the development of oral stereognosis in children and adults, where stereognosis is the ability to perceive and recognize the form of an object in the absence of visual and auditory information, by using tactile information. In oral stereognosis, the form of an object is recognized by exploring tactile information such as texture, size or spatial properties, in the oral cavity. This is usually evaluated by comparing the ability of children and adults to differentiate or identify small plastic objects in their mouths. Researchers have reported that oral sensory discrimination skills depend on age (McDonald and Aungst, 1967; Dette and Linke, 1982; Gisel and Schwob, 1988). McDonald and Aungst (1967) showed that 6- to 8-year-old children correctly matched half of the presented forms; 17- to 31-year-old adolescents and adults had perfect scores; and scores declined significantly with age among the 52- to 89-year-olds. Dette and Linke (1982) found similar results in 3- to 17-year-olds. The effect of age was also found in younger vs. older children. Kumin et al. (1984) showed that among 4- to 11-year-olds, the older children had significantly better oral stereognosis scores than younger children. Gisel and Schwob (1988) reported that 7- and 8-year-old children had better identification skills in an oral stereognosis experiment than 5- and 6-year-old children. Interestingly, only the 8-year-old children showed a learning effect, in that they got better scores as the experiment progressed.

To explain this age-related improvement in oral stereognosis, it was suggested that oral stereognosis maturity is achieved when the growth of the oral and facial structures is complete (McDonald and Aungst, 1967; Gisel and Schwob, 1988). This explanation is consistent with vocal tract growth data that shows that while major changes occur in the first 3 years of life (Vorperian et al., 1999), important growth of the pharyngeal region is observed between puberty and adulthood (Fitch and Giedd, 1999) and multidimensional maturity of the vocal tract is not reached until adulthood (Boë et al., 2007, 2008).

A few recent studies have suggested that there is a link between auditory and somatosensory information in multimodal integration.

Lametti et al. (2012) proposed that sensory preferences in the specification of speech motor goals could mediate responses to real-time manipulations, which would explain the important variability in compensatory behavior to an auditory

manipulation (Purcell and Munhall, 2006; Villacorta et al., 2007; MacDonald et al., 2010). They point out that one's own auditory feedback is not the only reliable source of speech monitoring and, in line with the internal speech model theory, that somatosensory feedback would also be considered in speech motor control. In agreement with this concept, Katseff et al. (2012) suggested that partial compensation in auditory manipulation of real-time speech could be because both auditory and somatosensory feedback system monitor speech motor control and therefore, the two systems are competing when large sensory manipulation affects only one of the sensory channels.

A recent study of speech auditory feedback perturbations in blind and sighted speakers supports the latter explanation. It showed that typically developing adults, whose somatosensory goals are narrowed by vision were more likely to tolerate large discrepancies between the expected and produced auditory outcome, whereas blind speakers, whose auditory goals had primacy over somatosensory ones, tolerated larger discrepancies between their expected and produced somatosensory feedback. In this sense, blind speakers were more inclined to adopt unusual articulatory positions to minimize divergences of their auditory goals (Trudeau-Fisette et al., 2017).

Researchers have also suggested that acoustic and somatosensory cues are integrated. As far as we know, Von Schiller (cited in Krueger, 1970; Jousmäki and Hari, 1998) was the first one to report that sound could modulate touch. Indeed, although he was mainly focused on the interaction between auditory and visual cues, he showed in his 1932s article that auditory stimuli, such as tones and noise bursts, could influence an object's physical perception. Since then, studies have shown how manipulations of acoustic frequencies or even changes in their prevalence can influence the tactile perception of objects, events, and skin deformation such as their perceived smoothness, occurrence, or magnitude (Krueger, 1970; Jousmäki and Hari, 1998; Guest et al., 2002; Hötting and Röder, 2004; Ito and Ostry, 2010). Multimodal integration was stronger when both perceptual sources were presented simultaneously (Jousmäki and Hari, 1998; Guest et al., 2002).

This interaction between auditory and tactile channels is also found in the opposite direction, in that somatosensory inputs can influence the perception of sounds. For example, Schürmann et al. (2004) showed that vibrotactile cues can influence the perception of sound loudness. Later, Gick and Derrick (2009) demonstrated that aerotactile inputs could modulate the perception of a consonant's oral property.

Somatosensory information coming from orofacial areas is somewhat different from those typically intended. Kinesthetic feedback usually refers to information retrieved from position, movement, and receptors in muscles and articulators (Proske and Gandevia, 2009). However, some of the orofacial regions involved in speech production movement are devoid of muscle proprioceptors. Therefore, the somatosensory information guiding our perception and production abilities likely also come from cutaneous mechanoreceptors (Johansson et al., 1988; Ito and Gomi, 2007; Ito and Ostry, 2010).

Although many studies have reported on the role of somatosensory information derived from orofacial movement in speech production (Tremblay et al., 2003; Nasir and Ostry, 2006; Ito and Ostry, 2010; Feng et al., 2011; Lametti et al., 2012), few studies have reported its role in speech perception.

Researchers recently investigated the contribution of somatosensory information on speech perception mechanisms. Ito et al. (2009) designed a bimodal perceptual task experiment where they asked participants to identify if the perceived target was the word "head" or "had." When the acoustic targets (all members of the "head/had" continuum) were perceived simultaneously to a skin manipulation recalling the oral articulatory gestures implicated in the production of the vowel /ɛ/, the identification rate of the target "head" was significantly improved. The researchers also tested different directions of the orofacial muscle manipulation and established that the observed effect was only found if the physical manipulation reflected a movement required in speech production (Ito et al., 2009).

Somatosensory information appears to even be involved in the processing of higher-level perceptual concepts (Ogane et al., 2017). In a similar perceptual task, participants were asked to identify if the perceived acoustic target was "l'affiche" (the poster) or "la fiche" (the form). The authors showed that the appropriate temporal positions of somatosensory skin manipulation in the stimulus word, simulating somatosensory inputs concerning the hyperarticulation of either the vowel /a/ or the vowel /i/, could affect the categorization of the lexical target.

Although further study would reinforce these findings, these experiments highlight the fact that the perception of linguistic inputs can be influenced by the manipulation of cutaneous receptors involved in speech motion (Ito et al., 2009, 2014; Ito and Ostry, 2010), and furthermore, attest of a strong link between auditory and somatosensory channels within the multimodal aspect of speech perception in adults.

The fact that sounds discrimination if facilitated when included in the infants' babbling register (Vihman, 1996) is surely part of the growing body of evidence that demonstrates how somatosensory information that is derived from speech movement also influences speech perception in young speakers (DePaolis et al., 2011; Bruderer et al., 2015; Werker, 2018). However, to our knowledge, only two studies have investigated how somatosensory feedback is involved in speech perception abilities in children (Yeung and Werker, 2013; Bruderer et al., 2015). In both studies, the researchers manipulated oral somatosensory feedback by constraining tongue or lip movement, thus forcing the adoption of a precise articulatory position. Although MSI continues to evolve until late childhood (Ross et al., 2011), these two experiments in toddlers shed light on how this phenomenon emerges.

In their 2013 article, Yeung and Werker (2013) reported that when 4- and 5-month-old infants were confronted with incongruent auditory and labial somatosensory cues, they were more likely to fix the visual demonstration corresponding to the vowel perceived through the auditory channel. In contrast, congruent auditory and somatosensory cues did not call for the need to add a corresponding visual representation of the perceived vowel.

Also using a looking-time procedure, Bruderer et al. (2015) focused on the role of language experience on the integration

of somatosensory information. They found that the ability of 6-month-old infants to discriminate between the non-native dental /d̪/ and the retroflex /ɖ/ Hindi consonant was influenced by the insertion of a teething toy. When the toddlers' tongue movements were restrained, they showed no evidence of phonetic contrast discrimination of tongue tip position. As shown by Ito et al. (2009), the effect of somatosensory cues was only observed if the perturbed articulator would have been involved in the production of the sound that was heard.

While these two studies mainly focused on perceptual discrimination rather than categorical representation of speech, they suggest that proprioceptive information resulting from static articulatory perturbation plays an important role in speech perception mechanisms in toddlers and that the phenomenon of multimodal integration in the perception-production speech model starts early in life. The authors suggested that, even at a very young age, babies can recognize that information can come from multiple sources and they react differently when the sensory sources are compatible. However, it is still unknown when children begin to integrate various sensory sources to treat them as a single sensory source.

In the current study, we aimed to investigate how dynamic somatosensory information from orofacial cutaneous receptors is integrated in speech processing in children compared to adults. Based on previous research, we hypothesized that: (1) when somatosensory inputs are presented simultaneously with auditory inputs, this affects their phonemic categorization; (2) auditory and somatosensory integration is stronger in adults than in children; and (3) MSI is facilitated when both types of sensory feedback are consistent.

## MATERIALS AND METHODS

### Participants

We recruited 15 young adults (aged 19–30), including eight females. We also recruited 21 children (aged 4–6) and after excluding seven children due to equipment malfunction (1), non-completion (2), or inability to understand the task (4), this left 14 children (aged 5–6) including 10 females, for the data analysis. Five- to six-year-old is a particularly interesting age window since children master all phonemes of their native language. However, they have not yet entered the fluent reading

stage, during which explicit teaching of reading has been shown to alter multimodal perceptual (Horlyck et al., 2012).

All participants were native speakers of Canadian French and were tested for pure-tone detection threshold using an adaptive method (DT < 25 dB HL at 250, 500, 1,000, 2,000, 4,000 and 8,000 Hz). None of the participants reported having speech or language impairments. The research protocol was approved by the Université du Québec à Montréal's Institutional Review Board (no 2015-05-4.2) and all participants (or the children's parents) gave written informed consent. The number of participants was limited due to the age of the children and the length of the task (3 different tasks were executed on the same day).
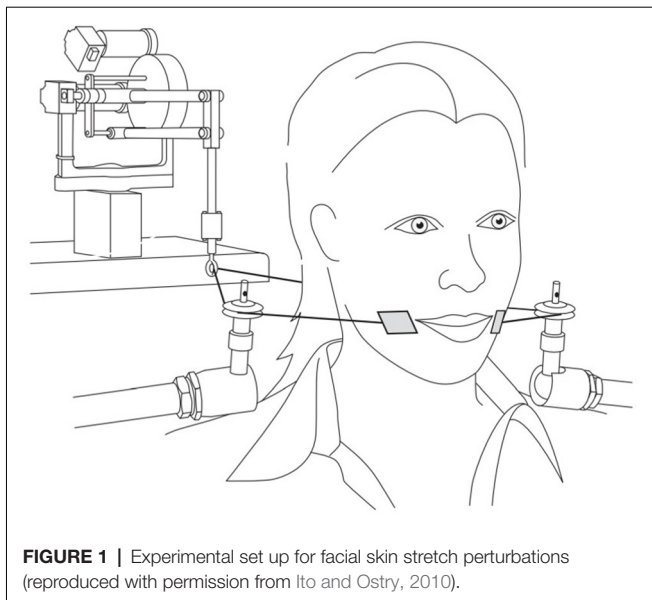
### Experimental Procedure

As in the task used by Ito et al. (2009), the participants were asked to identify the vowel they perceived and were asked to choose between /e/ and /ø/. Based on Ménard and Boe (2004), the auditory stimulus consisted of 10 members of a synthesized /e–ø/ continuum generated using the Maeda model (see **Table 1**). This continuum was created such that the first four formants were equally distributed from those corresponding to the natural endpoint tokens of /e/ and /ø/. To ensure that the children understood the difference between the two vocalic choices, the vowel /e/ was represented by an image of a fairy (/e/ as in fée) and the vowel /ø/ was represented by an image of a fire (/ø/ as in feu). Since, we wanted to minimize large head movements during the experiment, the children were asked to point out the image corresponding to their answers. Both images were placed in front of them at shoulder level, three feet away from each other on the horizontal plane. The adults were able to use the keyboard without looking at it and they used the right and left arrows to indicate their responses.

**Figure 1** shows the experimental set-up for the facial skin stretch perturbations. The participants were seated with their backs to a Phantom 1.0 device (SensAble Technologies) and they wore headphones (Sennheiser HD 380 pro). This small unit, composed of a robotic arm to which a wire is attached, allows for minor lateral skin manipulation at the side of the mouth, where small plastic tabs (2 mm × 3 mm), located on the ends of the wire, were placed with double-sided tape. The robotic arm was programed to ensure that when a four Newton

**TABLE 1 |** Formant and bandwidth values of the synthesized stimuli used in the perceptual task.

| | Formant values | | | | | Bandwidths values | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | F1 | F2 | F3 | F4 | F5 | B1 | B2 | B3 | B4 | B5 |
| Auditory stimuli | | | | | | | | | | |
| 1 | 364 | 1,922 | 2,509 | 3,550 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 2 | 364 | 1,892 | 2,469 | 3,500 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 3 | 364 | 1,862 | 2,429 | 3,450 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 4 | 364 | 1,832 | 2,389 | 3,400 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 5 | 364 | 1,802 | 2,349 | 3,350 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 6 | 364 | 1,772 | 2,309 | 3,300 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 7 | 364 | 1,742 | 2,269 | 3,250 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 8 | 364 | 1,712 | 2,229 | 3,200 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 9 | 364 | 1,682 | 2,189 | 3,150 | 4,000 | 48 | 55 | 60 | 50 | 100 |
| 10 | 364 | 1,652 | 2,149 | 3,100 | 4, 000 | 48 | 55 | 60 | 50 | 100 |

**FIGURE 1** | Experimental set up for facial skin stretch perturbations (reproduced with permission from Ito and Ostry, 2010).

flexion force was administered it led to a 10- to 15-mm lateral skin stretch.

When this facial skin stretch is applied at lateral to the oral angle in the backward direction as shown in the figure, it mimics the articulation associated with the production of the unrounded vowel /e/. Therefore, auditory and somatosensory feedback was either congruent (with /e/-like auditory inputs) or incongruent (with /ø/-like auditory inputs). As stated early, cutaneous receptors found in the within the labial area provides speech related kinesthetic information (Ito and Gomi, 2007). Since the skin manipulation was programed to be perceived at the same time as the auditory stimuli, it was possible to investigate the contribution of the somatosensory system to the perceptual processing of the speech targets.

The auditory stimuli were presented in 20 blocks of 10 trials each. Within each block, all members of the 10-step continuum were presented in a random order. For half of the trials, only the auditory stimulus was presented (unimodal condition). For the other half of the trials, a facial skin manipulation was also applied (bimodal condition). Alternate blocks of unimodal and bimodal conditions were presented to the participants. In total, 200 perceptual judgments were collected, 100 in the auditory-only condition and 100 in the combined auditory and skin-stretch condition.

## Data Analysis

For each participant, stimulus, and condition, we calculated the percentage of /e/ responses. The experiment was closely monitored, and the responses in trials where a short pause was requested by the participant were excluded from the analysis. In doing so, we sought to eliminate categorical judgments for which the participants were no longer in a position to properly respond to the task (fewer than 1.1% and 0.2% of all responses were excluded for children and adults, respectively). These perceptual scores were then fitted onto a logistic regression model (Probit model) to obtain psychometric functions from

which the labeling slopes and 50% crossover boundaries were computed. The value of the slope corresponds to the sharpness of the categorization (the lower the value, the more distinct the categorization), while the boundary value indicates the location of the categorical boundary between the two vowel targets (the higher the value, the more toward /ø/ the frontier). Using the lme4 package in R, we carried out a linear mixed-effects model (Baayen et al., 2008) for both the steepness of the slopes and the category boundaries in which group (adult or children) and condition (unimodal or bimodal) were specified as fixed factors and individual participant was defined as a random factor.

Each given answer (5,800 perceptual judgments collected from 29 participants) was fitted into a linear mixed-effects model where fixed factors included stimuli (the 10-step continuum), group (adult or children), and condition (unimodal or bimodal), and the random factor was the individual participant. The mean categorization of the first and last two stimuli was also compared. Once again, the averages of the given answers (116 mean perceptual judgments collected from 29 participants) were fitted into a linear mixed-effects model where the fixed variables included stimuli (head stimuli or tail stimuli), group (adult or children), and condition (unimodal or bimodal) and where the random variable was the individual participant. Finally, independent $t$-tests were carried out in order to compare variability in responses between both experimental groups and conditions. In both cases, Kolmogorov–Smirnov tests indicated that categorizations followed a normal distribution.

## RESULTS

The overall percentage of /e/ responses for each stimulus is shown in **Figure 2**. The data were averaged across speakers, within both groups. **Figure 3** displays the values for the labeling slope (distinctiveness of the vowels' categorization) and 50% crossover boundary (location of the categorical frontier) averaged across experimental conditions and groups. As can be seen in both figures, regardless of the experimental condition, the children had greater variations in overall responses compared to the adults, which was confirmed in an independent $t$-test ($t_{(38)} = 2.792, p < 0.01$).

## Psychometric Functions

### Labeling Slope Results

The linear mixed-effects model revealed a significant main effect of group on the steepness of the slope ($\chi^2_{(1)} = 23.549, p < 0.001$), indicating that there was more categorical perception in adults than in children (see **Figure 2**, black lines and **Figure 3**, left-hand part of the graph).

Although no effect of condition as a main effect was observed ($\chi^2_{(1)} = 3.618, p > 0.05$), a significant interaction between group and condition was found ($\chi^2_{(1)} = 4.956, p < 0.05$). *Post hoc* analysis revealed that in the bimodal condition the slope of the labeling function was more abrupt for the adults ($z = -3.153, p < 0.01$) but not for the children, suggesting that the skin stretch condition led to a more categorical identification of the stimuli in adults only.
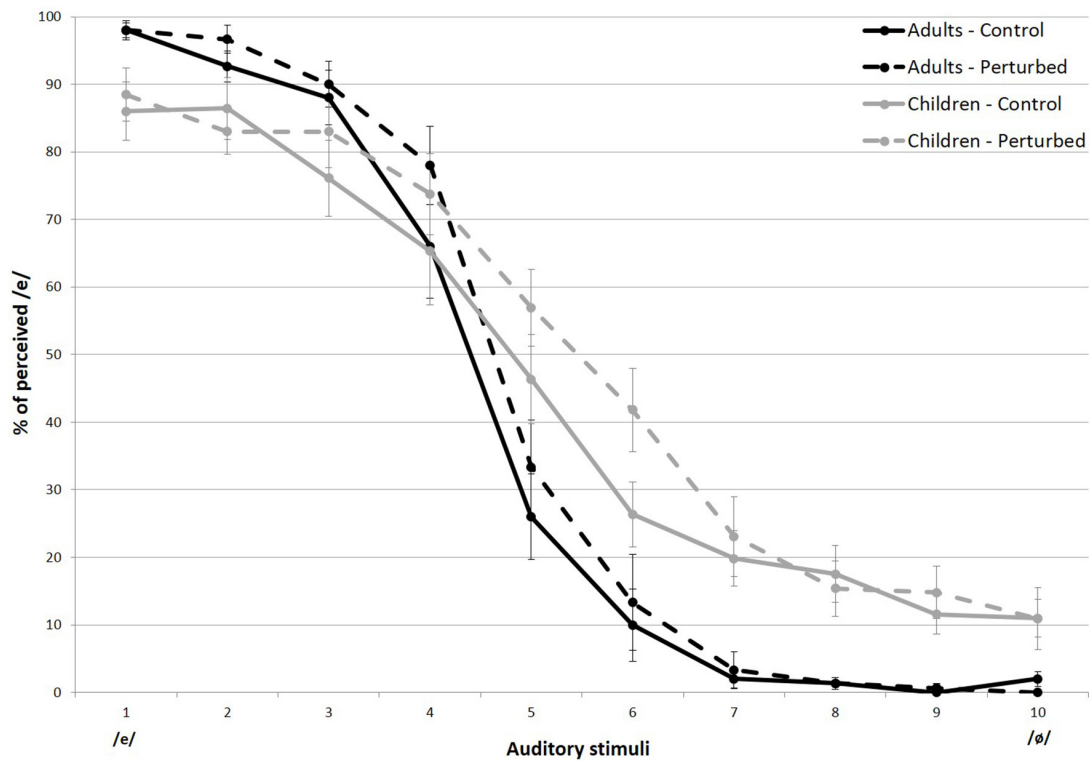
**FIGURE 2 |** Percent identification of the vowel [e] for stimuli on the [e–ø] continuum, in both experimental conditions, for both groups. Error bars indicate standard errors.

## The 50% Crossover Boundary Results

A linear mixed-effects model analysis carried out on the 50% crossover boundaries revealed a single main effect of condition ($\chi^2_{(1)}$ = 9.245, $p < 0.01$). For both groups, the skin stretch perturbation led to a displacement of the 50% crossover boundary. In the bimodal condition (A+SS), the boundary was located closer to /ø/ than in the unimodal condition (A). This result is consistent with the expected effect of the skin stretch perturbation; more stimuli were perceived as /e/ than /ø/. No effect of group, as a main effect or with condition was found. The results are presented in **Figure 2** and in **Figure 3**, in the right-hand part of the graphs.
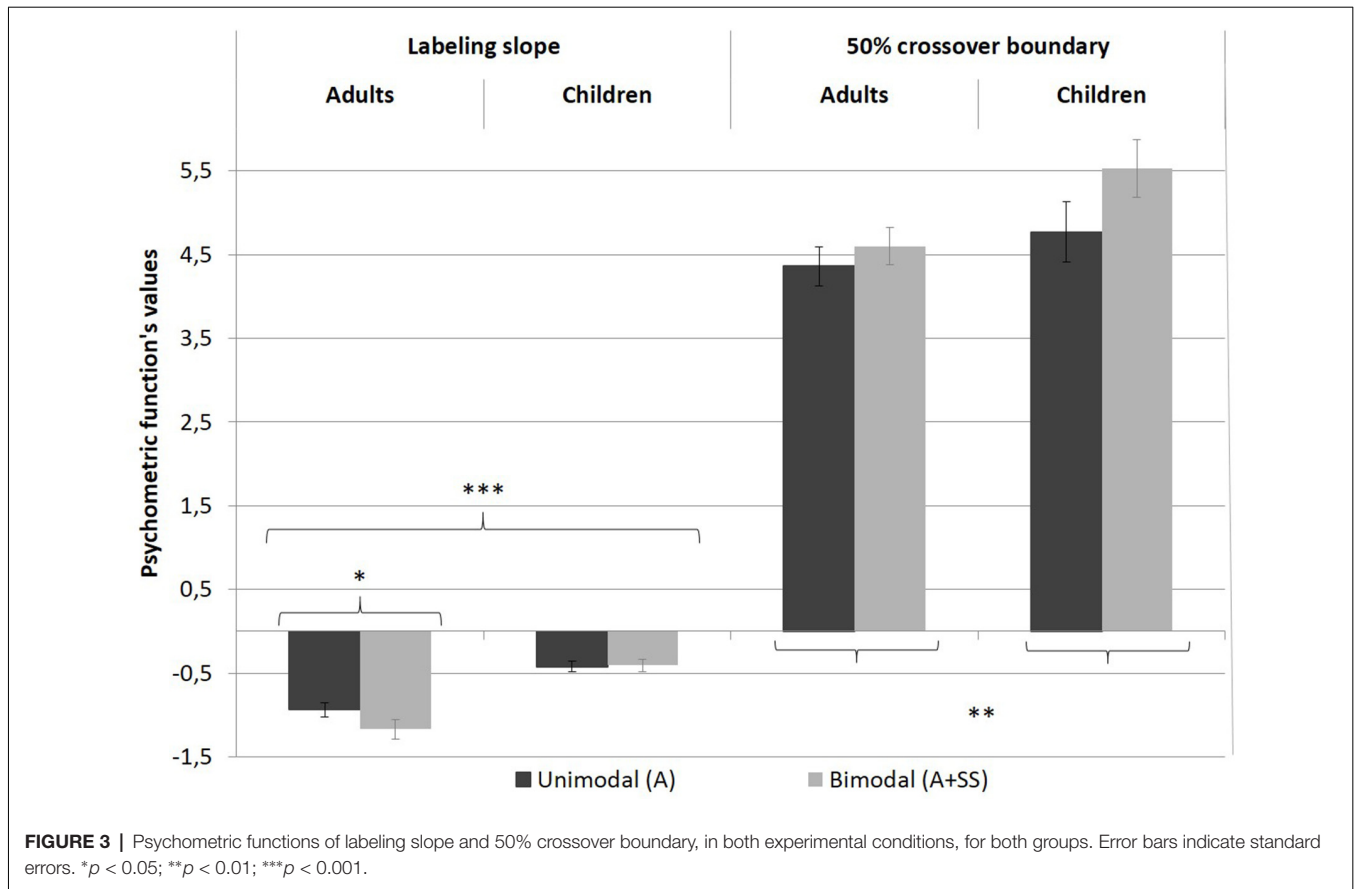
## Categorical Judgments

A linear mixed-effects model analysis performed on the categorical judgments revealed that in addition to the expected main effect of stimuli ($\chi^2_{(1)}$ = 3652.4, $p < 0.001$), there were significant effects of group ($\chi^2_{(1)}$ = 4.586, $p < 0.05$) and condition ($\chi^2_{(1)}$ = 15.736, $p < 0.001$), suggesting that children and adults did not categorize the stimuli in a similar manner and that both experimental conditions prompted different categorization. Moreover, a significant interaction of group and stimuli ($\chi^2_{(1)}$ = 144.52, $p < 0.001$) revealed that irrespectively of the experimental condition, some auditory stimuli were categorized differently by the two groups.

*Post hoc* tests revealed that whether a skin stretch manipulation was applied or not, stimulus 7 (A $z = -3.795$, $p < 0.1$ A+SS $z = -4.648$, $p < 0.01$), 8 (A $z = -3.445$, $p < 0.5$ A+SS $z = -3.544$, $p < 0.1$) and 9 (A $z = -3.179$, $p < 0.5$ A+SS $z = -4.347$, $p < 0.01$) were more systematically identified as /ø/ by the adults than by the children. While no other two-way interactions were found, a significant three-way interaction of group, condition, and stimuli was observed ($\chi^2_{(4)}$ = 117.26, $p < 0.001$) suggesting that, for some specific stimuli, the skin stretch condition affected the perceptual judgment of both groups in a different manner.

First, it was found that the skin stretch manipulation had a greater effect on stimulus 6, in children only ($z = -3.251$, $p < 0.5$). For this group, the skin stretch condition caused a 15.8% increase of /e/ labeling on stimulus 6. For the adults, the addition of somatosensory cues only led to a 3.3% increase in /e/ categorization.

Although less expected, the skin stretch manipulation also led to some perceptual changes at the endpoint of the auditory continuum. As shown in **Figure 2**, stimulus 2 ($z = 3.053$, $p < 0.5$) and stimulus 10 ($z = -3.734$, $p < 0.1$) were labeled differently by the two groups, but only in the bimodal condition. In fact, stimulus 2 (an /e/-like stimulus) was more likely to be identified as an /e/ by the adults in the experimental condition. In contrast, children were less inclined to label it so. As for stimulus 10 (an /ø/-like stimulus), the addition of somatosensory inputs

**FIGURE 3 |** Psychometric functions of labeling slope and 50% crossover boundary, in both experimental conditions, for both groups. Error bars indicate standard errors. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

decreased the correct identification rate in children only. In adults, although it barely affected their categorical judgments, the skin stretch manipulation mimicking the articulatory gestures of the vowel /e/ resulted in an increase of /ø/ labeling, as if it had a reverse effect.

Last, a comparison of mean categorizations of the first and last two stimuli revealed a main effect of stimuli ($\chi^2_{(1)} = 313.52$, $p < 0.001$) and a significant interaction of group and stimuli ($\chi^2_{(1)} = 36.260$, $p < 0.001$). More importantly, it also revealed a 3-way interaction of group, condition, and stimuli ($\chi^2_{(4)} = 37.474$, $p < 0.001$). *Post hoc* tests indicated that those endpoint stimuli of the continuum were categorized differently by the two groups, but only when a skin stretch manipulation was applied. In agreement with previous results, in the skin stretch condition, children labeled more /e/-like stimuli as /ø/ ($z = 3.434$, $p < 0.5$), and more /ø/-like stimuli as /e/ ($z = -4.139$, $p < 0.01$).

## DISCUSSION

This study aimed to investigate how auditory and somatosensory information is integrated in speech processing by school-aged children and adults, by testing three hypotheses.

As hypothesized, the overall perceptual categorization of the auditory stimuli was affected by the addition of somatosensory manipulations. The results for psychometric functions and categorical judgments revealed that auditory stimuli perceived

simultaneously with skin stretch manipulations were labeled differently than when they were perceived on their own. Sounds were more perceived as /e/ when they were accompanied by the proprioceptive modification.

The second hypothesis that auditory and somatosensory integration would be greater in adults than in children was also confirmed. As shown in **Figures 2**, **3**, orofacial manipulation affected the position of the 50% crossover boundary of both groups; when backward skin stretches were perceived simultaneously with the auditory stimulus, it increased its probability of being identified as an /e/. This impact of skin stretch manipulation on the value corresponding to the 50th percentile was also reported in Ito et al.'s (2009) experiment. However, bimodal presentation of auditory and somatosensory inputs affected the steepness of the slope in adults only. **Figure 2** also shows that adult participants were more likely to label /e/-like stimuli as /e/ in the bimodal condition. Since negligible changes were observed for /ø/-like stimuli, it led to a more categorical boundary between the two acoustic vocalic targets. This difference in the integration patterns between children and adults suggests that linkage of specific somatosensory inputs with a corresponding speech sound evolves with age.

The third hypothesis that MSI would be stronger when auditory and somatosensory information was congruent was confirmed in adults but not in children. Only adults' perception was facilitated when both sensory information was consistent.

In children, a decrease in the correct identification rate resulted from the bimodal presentation when auditory and proprioceptive inputs were compatible. Moreover, while adults seemed to not be affected by the /e/-like skin stretches when auditory stimuli were alongside the prototypical /ø/ vocalic sound (see **Figure 2**), children's categorization was influenced even when sensory channels were clearly contrasting, as if the bimodal presentation of vocalic targets blurred the children categorization abilities. Moreover, thought somatosensory information mostly affected specific stimuli in adult, it's effect in children was further distributed along the auditory continuum. These last observations support our second hypothesis that MSI is strongly defined in adults.

As many have suggested, MSI continues to develop during childhood (e.g., Ross et al., 2011; Dionne-Dostie et al., 2015). The fact that young children are influenced by somatosensory inputs in a different manner then adults could, therefore, be due to their underdeveloped MSI abilities. Related findings have been reported for audiovisual integration (McGurk and MacDonald, 1976; Massaro, 1984; Desjardins et al., 1997). It has also been demonstrated that the influence of visual articulators in audition is weaker in school-aged children than in adults.

In agreement with the concept that MSI continues to develop during childhood, the differences observed between the two groups of perceivers could also be explained by the fact that different sensory systems develop at different rates and in different ways. In that sense, it has also been found that school-aged children were not only less likely to perceive a perceptual illusion resulting from incongruent auditory and visual inputs, but they also had poorer results in the identification of unimodal visual targets (Massaro, 1984).

Studies of the development of somatosensory abilities also support this concept. As established earlier, oral sensory acuity continues to mature until adolescence (McDonald and Aungst, 1967; Dette and Linke, 1982; Holst-Wolf et al., 2016). The young participants who were 5–6 years of age in the current study may have had underdeveloped proprioceptive systems, which may have caused their less clearly defined categorization of bimodal presentations.

It is generally accepted that auditory discrimination is poorer and more variable in children than in adults (Buss et al., 2009; MacPherson and Akeroyd, 2014), and children's lower psychometric scores are often related to poorer attention (Moore et al., 2008).

MSI requires sustained attention, and researchers have suggested that poor psychometric scores in children might be related to an attentional bias between the recruited senses in children vs. adults (Spence and McDonald, 2004; Alsius et al., 2005; Barutchu et al., 2009). For example, Barutchu et al. (2009) observed a decline in multisensory facilitation when auditory inputs were presented with a reduced signal-to-noise ratio. They suggested that the increased level of difficulty in performing the audiovisual detection task under high noise condition may be responsible for the degraded integrative processes.

If this attention bias might explain some of the between-group performance differences found when /e/-like somatosensory inputs were presented with /ø/-like auditory inputs (high level

of difficulty), it would not justify differences between children and adults when the auditory and somatosensory channels agreed. The children showed decreased multisensory ability when both sensory inputs were compatible. Since difficulty level was reduced when multiple sensory sources were compatible, we should only have observed confusion in the children's categorization when auditory and somatosensory information was incongruent. According to the intersensory redundancy hypothesis, MSI should be improved when information from multiple sources is redundant. Indeed, Bahrick and Lickliter (2000) suggested that concordance of multiple signals would guide attention and even help learning (Barutchu et al., 2010). In the current study, this multisensory facilitation was only found in the adult participants.

This latter observation and the fact that no significant differences in variability were found across experimental conditions make it difficult to link the dissimilar patterns of MSI found between the two groups to an attention bias in children. However, finding a greater variability in MSI in children in both conditions, combined with their distinct psychometric and categorical scores provides support for the concept that perceptual systems in school-aged children are not yet fully shaped, which prevents them from attaining adult-like categorization scores.

As speech processing is multisensory and 5- to 6-year-olds have already experienced it, it is not surprising that some differences, even typical MSI ones, were found between the two experimental conditions in children. Since even very young children recognize that various speech sensory feedback can be compatible—or not (Patterson and Werker, 2003; Yeung and Werker, 2013; Bruderer et al., 2015; Werker, 2018), the different behavioral patterns observed in this study suggest that some form of multimodal processing exists in school-aged children, but complete maturation of the sensory systems is needed to achieve adult-like MSI.

## CONCLUSION

When somatosensory input was added to auditory stimuli, it affected the categorization of stimuli at the edge of the categorical boundary for both children and adults. However, while the oral skin stretch manipulation had a defining effect on phonemic categories in adults, it seemed to have a blurring effect in children, particularly on the prototypical auditory stimuli. Overall, our results suggest that since adults have fully developed sensory channels and more experiences in MSI, they have stronger auditory and somatosensory integration than children.

Although longitudinal observations are not possible, two supplementary experiments in these participants has been conducted to further investigate how MSI takes place in speech processing in school-aged children and adults. These focus on the role of visual and auditory feedback.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Comité institutionnel d'éthique de la recherche avec des êtres humaines (CIERH) de l'Université du Québec à Montréal [UQAM; Institutional review board of research ethics with humans of the Université of Québec in Montréal (UQAM)]" with written informed consent from all subjects or their parent (for minor). All subjects or their parent gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the "Comité institutionnel d'éthique de la recherche avec des êtres humaines (CIERH) de l'Université du Québec à Montréal [UQAM; Institutional review board of research ethics with humans of the Université of Québec in Montréal (UQAM)]".

## AUTHOR CONTRIBUTIONS

PT-F, TI and LM contributed to the conception and design of the study. PT-F collected data, organized the database and performed the statistical analysis (all under LM's guidance). PT-F wrote the first draft of the manuscript. PT-F and LM were involved in subsequent drafts of the manuscript. PT-F, TI and LM contributed to manuscript revision, read and approved the submitted version.

## FUNDING

## ACKNOWLEDGMENTS

## REFERENCES

Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Curr. Biol.* 15, 839–843. doi: 10.1016/j.cub.2005.03.046

Arabin, B. (2002). Music during pregnancy. *Ultrasound Obstet. Gynecol* 20, 425–430. doi: 10.1046/j.1469-0705.2002.00844.x

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Bahrick, L. E., and Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev. Psychol.* 36, 190–201. doi: 10.1037//0012-1649.36.2.190

Bahrick, L. E., and Lickliter, R. (2012). "The role of intersensory redundancy in early perceptual, cognitive and social development," in *Multisensory Development*, eds A. Bremner, D. J. Lewkowicz and C. Spence (Oxford: Oxford University Press), 183–205.

Barutchu, A., Crewther, D. P., and Crewther, S. G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Dev. Sci.* 12, 464–473. doi: 10.1111/j.1467-7687.2008.00782.x

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., and Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *J. Exp. Child Psychol.* 105, 38–50. doi: 10.1016/j.jecp.2009.08.005

Birch, H. G., and Lefford, A. (1963). Intersensory development in children. *Monogr. Soc. Res. Child Dev.* 28, 1–47. doi: 10.2307/1165681

Boë, L.-J., Granat, J., Badin, P., Autesserre, D., Pochic, D., Zga, N., et al. (2007). "Skull and vocal tract growth from newborn to adult," in *7th International Seminar on Speech Production, ISSP7* (Ubatuba, Brazil), 533–536. Available online at: https://hal.archives-ouvertes.fr/hal-00167610. Accessed March 15, 2019.

Boë, L.-J., Ménard, L., Serkhane, J., Birkholz, P., Kröger, B., Badin, P., et al. (2008). La croissance de l'instrument vocal: contrôle, modélisation, potentialités acoustiques et conséquences perceptives. *Revue Française de Linguistique Appliquée* 13, 59–80. Available online at: https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2008-2-page-59.htm. Accessed March 15, 2019.

Bower, T. G., Broughton, J. M., and Moore, M. K. (1970). The coordination of visual and tactual input in infants. *Percept. Psychophys.* 8, 51–53. doi: 10.3758/bf03208933

Bruderer, A. G., Danielson, D. K., Kandhadai, P., and Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proc. Natl. Acad. Sci. U S A* 112, 13531–13536. doi: 10.1073/pnas.1508631112

Burr, D., and Gori, M. (2011). "Multisensory integration develops late in humans," in *The Neural Bases of Multisensory Processes*, eds M. M. Murray and M. T. Wallace (Boca Raton, FL: CRC Press), 810.

Buss, E., Hall, J. W., and Grose, J. H. (2009). Psychometric functions for pure tone intensity discrimination: slope differences in school-aged children and adults. *J. Acoust. Soc. Am.* 125, 1050–1058. doi: 10.1121/1.3050273

Demany, L., McKenzie, B., and Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature* 266, 718–719. doi: 10.1038/266718a0

DePaolis, R. A., Vihman, M. M., and Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behav. Dev.* 34, 590–601. doi: 10.1016/j.infbeh.2011.06.005

Desjardins, R. N., Rogers, J., and Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *J. Exp. Child Psychol.* 66, 85–110. doi: 10.1006/jecp.1997.2379

Dette, M., and Linke, P. G. (1982). The development of oral and manual stereognosis in children from 3 to 10 years old (Die Entwicklung der oralen und manuellen Stereognose bei Kindem im Altervon 3 bis 10 Jahren). *Stomatologie* 32, 269–274.

Dionne-Dostie, E., Paquette, N., Lassonde, M., and Gallagher, A. (2015). Multisensory integration and child neurodevelopment. *Brain Sci.* 5, 32–57. doi: 10.3390/brainsci5010032

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306. Available online at: http://home.fau.edu/lewkowic/web/Eimas infant speech discrim Science 1971.pdf. Accessed February 1, 2019.

Feng, Y., Gracco, V. L., and Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *J. Neurophysiol.* 106, 667–679. doi: 10.1152/jn.00638.2010

Fitch, W. T., and Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106, 1511–1522. doi: 10.1121/1.427148

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J. Neurophysiol.* 88, 540–543. doi: 10.1152/jn.2002.88.1.540

Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502–504. doi: 10.1038/nature08572

Gisel, E. G., and Schwob, H. (1988). Oral form discrimination in normal 5-to 8-year-old children: an adjunct to an eating assessment. *Occup. Ther. J. Res.* 8, 195–209. doi: 10.1177/153944928800800403

Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694–698. doi: 10.1016/j.cub.2008.04.036

Guenther, F. H., and Perkell, J. S. (2004). "A neural model of speech production and its application to studies of the role of auditory feedback in speech," in *Speech Motor Control in Normal and Disordered Speech*, eds R. Maassen, R. D. Kent, H. Peters, P. H. van Lieshout and W. Hulstijn (Oxford: Oxford University Press), 29–49.

Guenther, F. H., and Vladusich, T. (2012). A neural theory of speech acquisition and production. *J. Neurolinguistics* 25, 408–422. doi: 10.1016/j.jneuroling.2009.08.006

Guest, S., Catmur, C., Lloyd, D., and Spence, C. (2002). Audiotactile interactions in roughness perception. *Exp. Brain Res.* 146, 161–171. doi: 10.1007/s00221-002-1164-z

Hatwell, Y. (1987). Motor and cognitive functions of the hand in infancy and childhood. *Int. J. Behav. Dev.* 10, 509–526. doi: 10.1177/016502548701000409

Hecht, D., and Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Exp. Brain Res.* 193, 307–314. doi: 10.1007/s00221-008-1626-z

Hillock, A. R., Powers, A. R., and Wallace, M. T. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia* 49, 461–467. doi: 10.1016/j.neuropsychologia.2010.11.041

Holst-Wolf, J. M., Yeh, I.-L., and Konczak, J. (2016). Development of proprioceptive acuity in typically developing children: normative data on forearm position sense. *Front. Hum. Neurosci* 10:436. doi: 10.3389/fnhum.2016.00436

Horlyck, S., Reid, A., and Burnham, D. (2012). The relationship between learning to read and language-specific speech perception: maturation versus experience. *Sci. Stud. Read.* 16, 218–239. doi: 10.1080/10888438.2010.546460

Hötting, K., and Röder, B. (2004). Hearing cheats touch, but less in congenitally blind than in sighted individuals. *Psychol. Sci.* 15, 60–64. doi: 10.1111/j.0963-7214.2004.01501010.x

Ito, T., and Gomi, H. (2007). Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in speech. *Neuroreport* 18, 907–910. doi: 10.1097/WNR.0b013e32810f2dfb

Ito, T., and Ostry, D. J. (2010). Somatosensory contribution to motor learning due to facial skin deformation. *J. Neurophysiol.* 104, 1230–1238. doi: 10.1152/jn.00199.2010

Ito, T., Gracco, V. L., and Ostry, D. J. (2014). Temporal factors affecting somatosensory-auditory interactions in speech processing. *Front. Psychol.* 5:1198. doi: 10.3389/fpsyg.2014.01198

Ito, T., Tiede, M., and Ostry, D. J. (2009). Somatosensory function in speech perception. *Proc. Natl. Acad. Sci. U S A* 106, 1245–1248. doi: 10.1073/pnas.0810063106

Johansson, R. S., Trulsson, M., Olsson, K. A., and And Abbs, J. H. (1988). Mechanoreceptive afferent activity in the infraorbital nerve in man during speech and chewing movements. *Exp. Brain Res.* 72, 209–214. doi: 10.1007/bf00248519

Jousmäki, V., and Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Curr. Biol.* 8:R190. doi: 10.1016/s0960-9822(98)70120-4

Katseff, S., Houde, J. F., and Johnson, K. (2012). Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? *Lang. Speech* 55, 295–308. doi: 10.1177/0023830911417802

Krakauer, J. W., Mazzoni, P., Ghazizadeh, A., Ravindran, R., and Shadmehr, R. (2006). Generalization of motor learning depends on the history of prior action. *PLoS Biol.* 4:e316. doi: 10.1371/journal.pbio.0040316

Krueger, L. E. (1970). David Katz's der aufbau der tastwelt (the world of touch): a synopsis. *Percept. Psychophys.* 7, 337–341. doi: 10.3758/BF03208659

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/bf03212211

Kuhl, P., and Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science* 218, 1138–1141. doi: 10.1126/science.7146899

Kumin, L. B., Saltysiak, E. B., Bell, K., Forget, K., Goodman, M. S., Goytisolo, M., et al. (1984). Relationships of oral stereognostic ability to age and sex of children. *Percept. Mot. Skills* 59, 123–126. doi: 10.2466/pms.1984.59.1.123

Lametti, D. R., Nasir, S. M., and Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *J. Neurosci.* 32, 9351–9358. doi: 10.1523/JNEUROSCI.0404-12.2012

Lecanuet, J.-P., Granier-Deferre, C., and Busnel, M.-C. (1995). "Human fetal auditory perception," in *Fetal Development: A Psychobiological Perspective*, eds J.-P. Lecanuet, W. P. Fifer, N. A. Krasnegor and W. P. Smotherman (Hillsdale, NJ: Lawrence Erlbaum Associates), 512.

Macaluso, E., Frith, C. D., and Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science.* 289, 1206–1208. doi: 10.1126/science.289.5482.1206

MacDonald, E. N., Goldberg, R., and Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *J. Acoust. Soc. Am.* 127, 1059–1068. doi: 10.1121/1.3278606

MacPherson, A., and Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. *Trends Hear.* 18:2331216514537722. doi: 10.1177/2331216514537722

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Dev.* 55, 1777–1788.

Massaro, D. W. (1999). Speech reading: illusion or window into pattern recognition. *Trends Cogn. Sci.* 3, 310–317. doi: 10.1016/S1364-6613(99)01360-1

McDonald, E. T., and Aungst, L. F. (1967). "Studies in oral sensorimotor function," in *Symposium on Oral Sensation and Perception*, ed. J. F. Bosma (Springfield, IL: Charles C. Thomas), 202–220.

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

McGurk, H., and Power, R. P. (1980). Intermodal coordination in young children: vision and touch. *Dev. Psychol.* 16, 679–680. doi: 10.1037/0012-1649.16.6.679

Ménard, L., and Boe, L. (2004). L'émergence du système phonologique chez l'enfant : l'apport de la modélisation articulatoire. *Cana. J. Linguist.* 49, 155–174. doi: 10.1017/S0008413100000785

Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., and Molholm, S. (2013). Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. *Neuroimage* 79, 19–29. doi: 10.1016/j.neuroimage.2013.04.060

Misceo, G. F., Hershberger, W. A., and Mancini, R. L. (1999). Haptic estimates of discordant visual-haptic size vary developmentally. *Percept. Psychophys.* 61, 608–614. doi: 10.3758/bf03205533

Mishra, J., Martinez, A., Sejnowski, T. J., and Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J. Neurosci.* 27, 4120–4131. doi: 10.1523/JNEUROSCI.4912-06.2007

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128. doi: 10.1016/s0926-6410(02)00066-6

Moore, D. R., Ferguson, M. A., Halliday, L. F., and Riley, A. (2008). Frequency discrimination in children: perception, learning and attention. *Hear. Res.* 238, 147–154. doi: 10.1016/j.heares.2007.11.013

Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of cue integration in human navigation. *Curr. Biol.* 18, 689–693. doi: 10.1016/j.cub.2008.04.021

Nasir, S. M., and Ostry, D. J. (2006). Somatosensory precision in speech production. *Curr. Biol.* 16, 1918–1923. doi: 10.1016/j.cub.2006.07.069

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., and Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Dev. Sci.* 9, 454–464. doi: 10.1111/j.1467-7687.2006.00512.x

Ogane, R., Schwartz, J.-L., and Ito, T. (2017). "Somatosensory information affects word segmentation and perception of lexical information," in *(Poster) Presented at the 2017 Society for the Neurobiology of Language conference in Baltimore, MD*. Available online at: https://hal.archives-ouvertes.fr/hal-01658527. Accessed March 05, 2019.

Patterson, M. L., and Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Dev. Sci.* 6, 191–196. doi: 10.1111/1467-7687.00271

Perrier, P. (1995). Control and representations in speech production. *ZAS Papers Lingustics* 40, 109–132. Available online at: https://hal.archives-ouvertes.fr/hal-00430387/document. Accessed March 05, 2019.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., and Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proc. Natl. Acad. Sci. U S A* 106, 10598–10602. doi: 10.1073/pnas.0904134106

Proske, U., and Gandevia, S. C. (2009). The kinaesthetic senses. *J. Physiol.* 587, 4139–4146. doi: 10.1113/jphysiol.2009.175372

Purcell, D. W., and Munhall, K. G. (2006). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J. Acoust. Soc. Am.* 120, 966–977. doi: 10.1121/1.2217714

Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., et al. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur. J. Neurosci* 31, 1772–1782. doi: 10.1111/j.1460-9568.2010.07213.x

Rentschler, I., Jüttner, M., Osman, E., Müller, A., and Caelli, T. (2004). Development of configural 3D object recognition. *Behav. Brain Res.* 149, 107–111. doi: 10.1016/s0166-4328(03)00194-3

Robert-Ribes, J., Schwartz, J.-L., Lallouache, T., and Escudier, P. (1998). Complementarity and synergy in bimodal speech: auditory, visual and audio-visual identification of French oral vowels in noise. *J. Acoust. Soc. Am.* 103, 3677–3689. doi: 10.1121/1.423069

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-amour, D., and Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *Eur. J. Neurosci.* 33, 2329–2337. doi: 10.1111/j.1460-9568.2011.07685.x

Schürmann, M., Caetano, G., Jousmäki, V., and Hari, R. (2004). Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. *J. Acoust. Soc. Am.* 115, 830–832. doi: 10.1121/1.1639909

Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399. doi: 10.1093/cercor/bhl147

Spence, C., and McDonald, J. (2004). "The cross-modal consequences of the exogenous spatial orienting of attention," in *The Handbook of Multisensory Processes*, eds G. A. Calvert, C. Spence and B. E. Stein (Cambridge, MA: MIT Press), 3–25.

Stein, B. E., London, N., Wilkinson, L. K., and Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J. Cogn. Neurosci.* 8, 497–506. doi: 10.1162/jocn.1996.8.6.497

Stein, B. E., and Meredith, M. A. (1993). *The Merging of the Senses.* Cambridge, MA: MIT Press.

Stein, B. E., Stanford, T. R., and Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nat. Rev. Neurosci.* 15, 520–535. doi: 10.1038/nrn3742

Streri, A., and Gentaz, E. (2004). Cross-modal recognition of shape from hand to eyes and handedness in human newborns. *Neuropsychologia* 42, 1365–1369. doi: 10.1016/j.neuropsychologia.2004.02.012

Tourville, J. A., and Guenther, F. H. (2011). The DIVA model: a neural theory of speech acquisition and production. *Lang. Cogn. Process.* 26, 952–981. doi: 10.1080/01690960903498424

Tremblay, S., Shiller, D. M., and Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature* 423, 866–869. doi: 10.1038/nature01710

Trudeau-Fisette, P., Tiede, M., and Ménard, L. (2017). Compensations to auditory feedback perturbations in congenitally blind and sighted speakers: acoustic and articulatory data. *PLoS One* 12:e0180300. doi: 10.1371/journal.pone.0180300

Turgeon, C. (2011). Mesure du développement de la capacité de discrimination auditive et visuelle chez des personnes malentendantes porteuses d'un implant cochléaire. Available online at: https://papyrus.bib.umontreal.ca/xmlui/handle/1866/6091. Accessed February 01, 2019.

Vihman, M. M. (1996). *Phonological Development: The Origins of Language in the Child.* Oxford, England: Blackwell.

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J. Acoust. Soc. Am.* 122, 2306–2319. doi: 10.1121/1.2773966

Vorperian, H. K., Kent, R. D., Gentry, L. R., and Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: preliminary results. *Int. J. Pediatr. Otorhinolaryngol.* 49, 197–206. doi: 10.1016/s0165-5876(99)00208-6

Walker-Andrews, A. (1994). "Taxonomy for intermodal relations," in *The Development of Intersensory Perception: Comparative Perspectives*, eds D. J. Lewkowicz and R. Lickliter (Hillsdale, NJ: Lawrence Erlbaum Associates), 39–56.

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Appl. Psycholinguist* 39, 703–728. doi: 10.1017/s014271641800022x

Yeung, H. H., and Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychol. Sci.* 24, 603–612. doi: 10.1177/0956797612458802

Yu, L., Rowland, B. A., and Stein, B. E. (2010). Initiating the development of multisensory integration by manipulating sensory experience,. *J. Neurosci.* 30, 4904–4913. doi: 10.1523/JNEUROSCI.5575-09.2010