Audio-aerotactile integration in infant speech perception[*]

Megan Keough
University of British Columbia

Abstract: Recent work has shown that adult native English speakers can integrate aerotactile information during speech perception. Moreover, this airflow cue both enhances and interferes with accurate speech perception. However, while considerable work has investigated the developmental trajectory of audiovisual speech perception, little is known about how and when audio-aerotactile integration arises. This paper reports on an experiment to test the hypothesis that this ability requires speech production experience. 12 preverbal infants took part in an Alternating/Non-Alternating Sound Presentation Task (Best & Jones, 1998; Yeung & Werker, 2009). They were presented with stimulus streams containing sequences of /ba/ and/or /pa/ syllables. Infants felt synchronous, gentle puffs of air during some of the tokens and it was hypothesized that the presence of the airflow on the unaspirated /ba/ tokens would cause the infants to treat the tokens as more /pa/-like. This would in turn influence their perception of the stimulus streams as alternating or non-alternating. Looking time data indicate that the infants integrated the multisensory cues, looking longer during trials that would only be alternating if the infant was processing the audio and aerotactile cues as part of the same speech event. The results demonstrate that the ability to integrate these cross-sensory cues does not arise as a result of speech production experience.

Keywords: multimodal speech perception, multisensory integration, aspiration, infant speech perception

1 Introduction

Though speech perception research often places primary focus on the auditory signal, a growing body of literature provides evidence for the multi-sensory nature of speech. Until fairly recently, much of this work focused specifically on how the integration of auditory and visual speech cues affect speech perception (e.g., McGurk & MacDonald, 1976). However, over the last twenty years, researchers have turned their attention to the role that the tactile modality plays. Some research has focused on somatosensory input that speakers receive from their own articulators (Ito, Tiede & Ostry, 2009) or tactile cues felt when placing a hand on their interlocutor's face (Gick, Jóhannsdóttir, Gibraiel, & Mühlbauer 2008). Others have looked at the aero-tactile or airflow input perceivers experience during the production of aspirated stops (Gick & Derrick, 2009; Goldenburg, Tiede, Whalen, 2015; Bicevskis, Derrick, & Gick, 2016) and fricatives (Derrick, O'Beirne, De Rybel, & Hay, 2014). In their seminal 2009 study, Gick and Derrick showed that aerotactile cues can both enhance and interfere with speech perception in much the

[*] A special thank you to the members of both the Interdisciplinary Speech Research Lab and the Infant Studies Centre for their assistance and support throughout. In particular, I wish to acknowledge Padmapriya Kandhadai and Murray Schellenberg for their help with the equipment set up and Savannah Nijeboer for her endless patience while I was learning to navigate the lab. Additional thanks to Donald Derrick for his assistance with the stats. This research was funded by a grant through the National Institutes of Health.

UBC Qualifying Papers 4 (2015-2017).
Emily Sadlier-Brown, Oksana Tkachman and Avery Ozburn (eds.). Vancouver: UBCWPL, 2019.

same way that visual information does. The authors asked native English-speaking participants to discriminate between pairs of tokens that contrasted in aspiration (/pa/ vs. /ba/ and /ta/ vs. /da/) in difficult listening conditions. When the syllables were accompanied by silent, naturalistic puffs of air applied to the participants' skin, participants were significantly more likely to perceive the syllable as aspirated (i.e., /pa/ or /ta/) *even if the air puff occurred with the unaspirated token.* This effect of aspiration has also been replicated in the visual-aerotactile domain (Biceskis et al., 2016), with fricatives (Derrick et al., 2014), and using a voicing continuum rather than voiced and voiceless exemplars (Goldenberg, Tiede, Whalen, 2015). The current study focuses on a question that then arises regarding aero-tactile integration: how and when does the ability to integrate this cue emerge? The current study seeks to chip away at this undoubtedly large research question by asking if this ability requires experience as a language producer to emerge.

Though no direct evidence exists to suggest that production experience plays a central role in the emergence of audio-aerotactile integration, it makes some intuitive sense that feeling one's airflow while simultaneously feeling one's articulators and hearing the acoustic output could provide a way to build a multimodal representation of sounds. Moreover, some computational models of speech production (e.g., the DIVA model Guenther, 1994; Guenther, 1995; Tourville & Guenther, 2011), propose that production experience during the initial babbling stage is what allows infants to create mappings between articulatory movements and their acoustic and sensory consequences. In the DIVA model, for example, the authors argue that babbling makes important contributions to sound acquisition because during this stage, infants create two mappings: one is phonetic-to-orosensory, and the other is orosensory-to-articulatory. Essentially, the first mapping links a sound with the vocal tract target that produces it, while the second mapping links the vocal tract movements with the motor commands needed to produce them. In this model, then, production experience plays a central role in linking acoustic outputs and the vocal tract configurations or movements that produce them.

An additional line of evidence that suggests production experience may affect our perception of sounds comes from research on speech perception in disordered populations: for example, in children with phonological disorders, difficulty producing a sound negatively affects their ability to perceive it. In other words, if children don't have experience accurately producing a sound, they have a harder time identifying it. As explained in Byun (2012), the misarticulations of a child who has a phonological disorder become a large part of the child's input. If the child accepts these incorrect productions as instances of a target phoneme, it could shift the boundaries of that phoneme and thus make it more difficult for the child to perceive. Furthermore, there is some evidence that production training can improve perceptual sensitivity. Shuster (1998) found that for children who with disordered production and perception of /ɹ/, therapy targeting their production of the sound resulted in significantly improved performance in a sound judgement task.

If the above suppositions hold, we might predict that 6-8 month old infants would not have the relevant experience needed to integrate. To date, no research has investigated aero-tactile integration in infant speech perception. However, there is evidence in the literature that infants can integrate other sensory cues well before they begin speaking. In the visual domain, for example, infants can map an acoustic signal to the face that matches the sound they are hearing (e.g., Kuhl & Metzhoff, 1982; Patterson & Werker, 1999; Patterson & Werker, 2003). McGurk-like effects have also been reported in infants as young as five months (Rosenblum, Schmuckler, & Johnson, 1997), though the effect seems to be weaker than in adults. Moreover, the literature suggests that infants can make use of tactile information during speech perception. For example, infants can be influenced by somatosensory feedback from their own oral tract while they are listening to speech sounds (Yeung & Werker, 2013; Bruderer, Danielson, Kandhadai, & Werker, 2015). All of this offers indirect support for the possibility of audio-aerotactile integration well before infants become producers of their native language.

The current study seeks to address the question regarding the developmental trajectory of audio-aerotactile integration with an experiment testing 6.5-8 month-old English-acquiring infants on the ability to use aero-tactile cues to distinguish between /pa/ and /ba/ in difficult listening conditions. Building on the fact that infants have been shown to distinguish between /p/ and /b/ in normal listening conditions at 6-8 months of age (e.g., Eimas, Siqueland, Juscyk, & Vigorito, 1971; though see Burns, Yoshida, Hill, & Werker, 2007 for evidence that infants may not discriminate the English boundary until after 8 months), the current study aims to see whether infants can recruit an additional cue (in this case, gentle puffs of air on the neck) to help them discriminate ambiguous stimuli. While infants at this age may have begun producing bilabial stops during babbling, research suggests that these stops are most likely voiceless unaspirated sounds with a short voicing lag. In fact, there is evidence to suggest that VOT may not be produced categorically until closer to two or even three years of age (Hitchcock & Koenig, 2013). Thus, it is unlikely that the 6-8 month-old infants in the current study would have experience feeling airflow across their own lips while producing [p^h]. Given this developmental trajectory, a production-based hypothesis predicts that the preverbal infants tested in the current study would not have access to airflow as a cue during discrimination because they do not have experience producing both sounds. If such a hypothesis is false, the infants should be able to use this airflow to discriminate between aspirated and unaspirated sounds. That is, they will treat unaspirated tokens (i.e., /ba/) accompanied by a puff of air as more like an aspirated token (i.e., /pa/). For the aspirated tokens, the prediction is less clear. The airflow on these syllables would theoretically serve as a redundant cue to information already present in the acoustic signal (i.e., the aperiodic noise of the aspiration). Indeed, for adults, the airflow increased accurate identification in Gick and Derrick (2009). Based on this result, then, we would predict that the infants will not treat a /pa/ accompanied by an air puff as equivalent to a plain /pa/. However, given that infants in this age range are still in the process of narrowing their native phonetic categories, they may not be judging the tokens on the basis of language-specific phonetic distinctions. If this is the case, then the infants may treat a /pa/ accompanied by a puff as a separate category altogether. Regardless, if we find that the infants' perception is *not* influenced by the airflow, this provides evidence that production experience plays an important role in audio-aerotactile integration. It would also raise the question of why some multisensory integration in speech perception requires production experience while others do not (e.g., audio-visual speech perception). On the other hand, if the infants *do* treat the unaspirated tokens as aspirated when they feel the puff of air, a production-based hypothesis would not be supported and other mechanisms must be proposed to explain the emergence of audio-aerotactile integration.

2 Methods

In the current study, the infants took part in a modified alternating/non-alternating sound presentation task (Best & Jones, 1998; Yeung & Werker, 2009). In this type of paradigm, infants are exposed to two different types of trials: an alternating trial, in which repetitions of two different sounds are presented (e.g., /ba/ and /pa/), and a non-alternating trial, in which repetitions of identical sounds are heard (e.g., /pa/ and /pa/). Often this paradigm employs a familiarization phase. However, the aim of the study was to test the infant's baseline ability to use aero-tactile information to discriminate between two sounds, rather than their ability to learn to use the cue.

This is crucial if the question at hand concerns whether the infants are currently able to integrate aero-tactile information during speech perception and not whether they can be taught to use it. Because of this, the choice was made not to employ a familiarization phase. Instead, the infants only experienced a series of test trials, as in Bruderer et al. (2015). In alternating/non-alternating paradigms, infants are assumed to have discriminated if they look longer to one type of trial than the other. Generally, in experiments without a familiarization phase, infants will look longer to an alternating stimulus. Therefore, in the current study, the infants are predicted to look longer to trials that they experience as alternating. As will be discussed further in Section 2.3 below, which trials they experience as alternating will depend on whether the infants are integrating the aero-tactile information.

2.1 Participants

12 English-acquiring infants (6 female; mean age = 7;11 range = 6;15-7;30) were recruited from a database of families who had been approached at a local maternity hospital shortly after birth and had indicated their interest in participating in studies. As measured through parent reporting, all infants were exposed to a minimum of 80% English and had not been diagnosed with any developmental disorders. Data from an additional nine infants were not included due to fussiness (n=6) and equipment error (n=2). Finally, data from one infant was excluded because the parents reported an undiagnosed lazy eye and it was very difficult to determine with any certainty whether the infant was looking at the screen. Before beginning the session, caregivers were informed about the study procedure and gave written consent for participation. At the end of the session, infants were given a t-shirt and a certificate as a token of appreciation for participating.

2.2 Apparatus and set up

Following Gick and Derrick (2009), an air compressor attached to a solenoid valve in a switchbox comprised the airflow device. The air puffs were delivered at ~6 p.s.i. through 1/4-inch vinyl tubing that passed through a cable port from the observation room to the study room. The tube then attached to the front of a flexible plastic bib around the infant's neck. This kept the mouth of the tube a constant 7 cm from the infant's neck and ensured that the airflow hit the infant's neck each time. The bib and tube were covered with fabric to keep the infant from grabbing or move the tubing (see Figure 1 below). Moreover, a custom sound-attenuating cloak attached to the high chair ran from the floor to just under the infant's chin. In effect, this created a separate acoustic space in which the airflow occurred, thus ensuring that the infants only experienced the airflow as a tactile sensation. Infants were excluded from analysis if at any point during the experiment the bib and tube came out from under the cloak as it could no longer be guaranteed that the infant was not hearing the air puff.

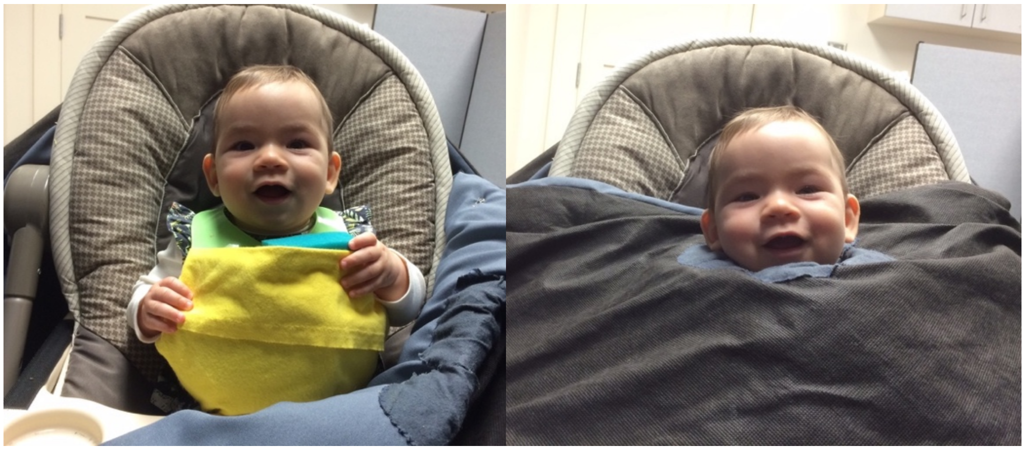

Figure 1: Photographs of the bib worn around the baby's neck and of the sound-attenuating smock. The vinyl tubing that delivered the airflow was located underneath the fabric and attached to the front of the bib. The tubing was then curved to aim the airflow at the baby's neck from a distance of 7 cm.

2.3 Stimuli

The auditory stimuli for each trial were created in a sound editing program (Audacity Team, 2016) by concatenating naturally produced tokens of /ba/ and /pa/ (six tokens each) that had been produced by a male native English speaker for the original Gick and Derrick (2009) study. The /ba/ tokens were phonetically voiceless unaspirated stops with VOTs less than 10 ms. The /pa/ tokens were phonetically voiceless aspirated stops with an average VOT of 60 ms. Twelve 20-second stimuli streams were created: six non-alternating (NonAlt) stimuli streams contained 12 presentations each of either /pa/ or /ba/ tokens at an ISI of 750 ms, and six alternating (Alt) stimuli streams contained six presentations each of /pa/ and /ba/ tokens at the same ISI. The tokens were placed in the right channel of a stereo track which was subsequently embedded in pink noise at +2 SNR. The noise was included both to reduce the ability to discriminate a native contrast they have been shown to be able to discriminate[1] and thus reduce the risk of ceiling effects, and to mask in part the noise of the airflow. In the left channel, 50-ms sine waves generated at a frequency of 10kHz triggered the release of the airflow. The waves were time-aligned with the syllables such that, after adjusting for system latency, the air puff exited the tube at the same time as the stop burst 50-ms (see Figure 2). This resulted in around 65 ms of aspiration and was done to mimic the natural timing of aspiration during English stops.

[1] Again, this is perhaps an oversimplification. As discussed above, researchers have found mixed results across task types.

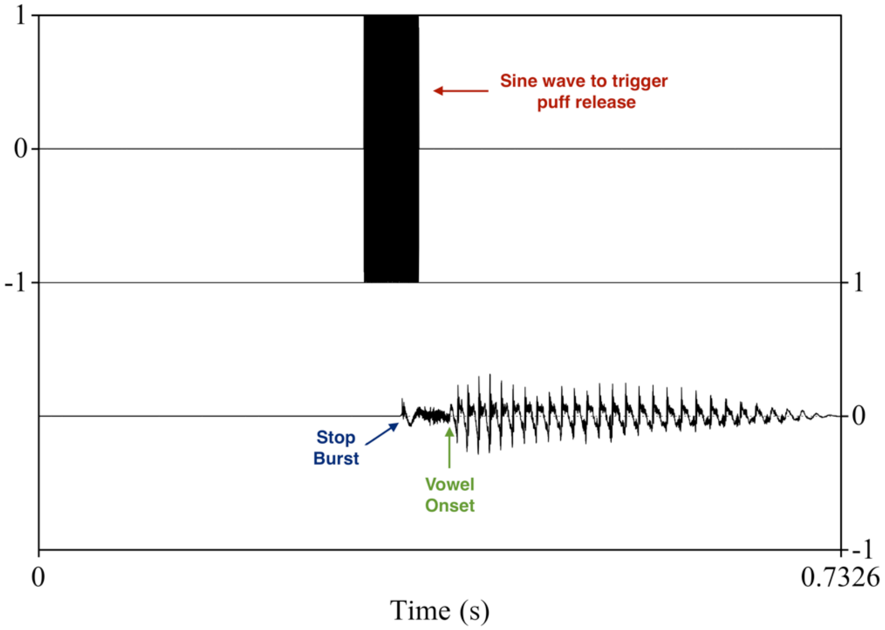

Figure 2: Placement of the sine wave relative to the stop burst and vowel onset. The 50ms tone was shifted an additional 20ms earlier to account for system latency.

In total, four stimuli stream types were created. Every other syllable was accompanied by a puff such that, as far as the tactile modality is concerned, all trials were alternating. As discussed above, the crucial difference across trials, then, was whether the tactile stimulus reinforced an existing phonological distinction or interfered with it, thereby influencing the infant's perception of whether the trial stimulus alternated. Table 1 below shows the four types of stimuli streams created, the trial type (Alt or NonAlt) predicted if the infants are only processing the acoustic signal, and the trial type (Alt or NonAlt) predicted if the infants are integrating the aerotactile stimulus as part of the speech event.

Table 1: The table below describes the four types of stimuli streams that the infants were presented with, as well as the predicted percept (i.e., an alternating or a non-alternating trial) depending on whether the infants integrate the aerotactile stimulus.

Stimuli stream	Predicted Trial Type Without Integration	Predicted Trial Type With Integration
paPuff + pa	NonAlt	NonAlt
paPuff + ba	Alt	Alt
pa + baPuff	Alt	NonAlt
ba + baPuff	NonAlt	Alt

2.4 Procedure

The infants were tested in a quiet, dimly lit room while seated in a high chair in front of a computer monitor that was positioned in the center of a black curtain. Caregivers were seated in a chair next to the infants. They were told not to point or talk during the session, though non-verbal reassurance, such as nodding or smiling, was encouraged to keep the infants calm. Caregivers also listened to music over headphones during the session to ensure they didn't unconsciously influence their child's reactions. A closed circuit camera was used to record the infant's face through a slit in the curtain directly below the computer monitor. From another room, an experimenter monitored the infant's face through the video display and controlled the stimulus presentation using computer software (Cohen, MacWhinney, Flatt, & Provost, 1993). The auditory stimuli were presented free field at ~65dB over a speaker located behind the black curtain.

The study began with a silent, colorful animation to attract the infant's attention. This was followed by a silent checkerboard trial to give the infants an opportunity to look at the novel stimulus before the test trials begin. Before the onset of each trial, the infant's attention was drawn to the monitor by a spinning waterwheel. Once the infant was looking at the screen, the test trial began and a red and black checkerboard and the stimuli stream were presented simultaneously. When the trial finished, the same waterwheel animation reappeared to return the infant's attention to the screen. The study consisted of 16 trials over two blocks with a 10 second animated video as a break in between blocks. The two blocks were identical except in the ordering of stimulus presentation. In each block, the infants were presented with four Alt trials and four NonAlt trials (in which the Alt or NonAlt designation reflects the predicted trial type if the infants are integrating the aero-tactile information as outlined in Table 1 above), with every other trial being alternating. The order of the first stimulus was counterbalanced across infants, such that half of the infants experienced an alternating stimuli stream first and half experienced a non-alternating stream. All infants experienced all stimuli stream types. The order of presentation for the aerotactile stimulus was also counterbalanced. As mentioned previously, an air puff was present on every other syllable in each trial. To control for potential order effects, half of the infants were presented with the air puff on the odd (first, third, fifth, etc) syllables, while the other half were presented with the air puff on the even (second, fourth, sixth, etc) syllables. In both counterbalancing cases, the order of presentation was then reversed for the second block.

The video recordings were converted to Quicktime movies. The looking time to test trials was coded offline frame by frame by the author. The trials were coded without audio so that the coder was blind to which type of trial the infant was experiencing. Total looking time to the checkerboard served as the dependent measure.

3 Analysis and Results

Looking time data for each infant were analyzed across 4 trials of each stimulus stream for a total of 16 trials. As Table 2 shows, the infants looked longer to the stimulus streams paPuff + ba and ba + baPuff, which were the two predicted alternating streams. Mean looking times were shortest for the pa + baPuff tokens, which infants would only treat as non-alternating if they are integrating the airflow information.

Table 2: Mean looking times and standard deviations for each stimulus stream type.

Stimulus Stream	**Mean looking time**	**Standard Error**
paPuff + pa	8.944833	0.71
paPuff + ba	9.859850	0.82
pa + baPuff	8.454742	0.64
ba + baPuff	9.247835	0.64

To test for statistical significance in looking time differences, a linear mixed-effects model was computed using the lmer4 package (Bates et al., 2015) in R (R Core Team, 2014) to predict looking times given the fixed effects of Stimulus Stream (paPuff + pa, paPuff + ba, pa + baPuff, ba +baPuff) and Trial Number, with a random effect of Subject, and a by-Subject random slope for Stimulus Stream and Trial Number.

3.1 Results

A significant effect of Trial Number (β = -0.44, *SE* = 0.12, *t* = -3.59, p < 0.001) emerged, such that the infants' looking times significantly decreased over the course of the session as is generally expected. In discussing the stimuli streams, I will return to the predictions outlined in Section 2.3 above. Recall that the overall prediction was that the presence of the aerotactile cue, or puff, would cause the infants to treat /ba/ syllables as more /pa/-like because they would incorporate the airflow during perception and show an aspiration effect like that seen in Gick and Derrick (2009). Given this prediction, the presence of the puff of air was predicted to shift the perceived nature of the trial (i.e., alternating or non-alternating) depending on whether the puff accompanied a sound that naturally produces a burst of air. In other words, the puff of air would interfere with the infant’s ability to discriminate between aspirated and unaspirated tokens only when the airflow occurred with a /ba/. It is also important to keep in mind that none of the stimulus streams was truly non-alternating because airflow was present on every other token. Thus, there is no real “baseline” or control against which to compare the other trials. A more useful approach, and the one used here, is to compare the stimulus streams to each other. In this way, we can test the prediction that infants will treat some stimulus types as more alternating than others (and therefore look longer to them) given the placement of the airflow. The stimulus stream predictions from Table 1 above are repeated here in Table 3 for ease of comparison.

Table 3: The table below describes the four types of stimuli streams that the infants were presented with, as well as the way the infants are predicted to perceive the trial (i.e., an alternating or a non-alternating trial) depending on whether the infants integrate the aerotactile stimulus.

Stimuli stream	**Predicted Trial Type Without Integration**	**Predicted Trial Type With Integration**
paPuff + pa	NonAlt	NonAlt
paPuff + ba	Alt	Alt
pa + baPuff	Alt	NonAlt
ba + baPuff	NonAlt	Alt

3.1.1 ba + paPuff vs. baPuff + pa

The first prediction was that infants would treat trials in which they were presented with paPuff + ba as more alternating, and thus look longer to them, than trials in which they were presented with pa + baPuff. This prediction is based on the assumption that the presence of the airflow in pa + baPuff trials would make the /ba/ seem more /pa/-like—and the trial less alternating—if infants are integrating the aerotactile information as part of the speech event. Indeed, this seems to be the case (see Figure 3). The infants looked significantly longer to ba + paPuff trials than they did to baPuff + pa trials (β = 1.66, *SE* = 0.64, *t* = 2.61, p = 0.03). Crucially, this result only matches the prediction with integration. If the infants were only paying attention to the acoustic information, then they would have been expected to treat the two trial types as roughly equivalent (i.e., both /ba/ + /pa/) and thus show no significant difference in looking times.

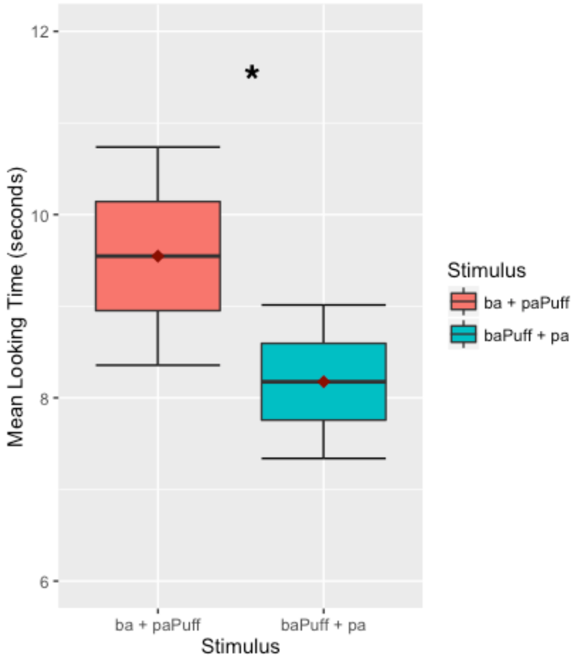

Figure 3: The above figure compares mean looking times for the two stimulus streams ba+ paPuff and baPuff + pa. As can be seen in the figure, the infants looked significantly longer to the ba + paPuff stimulus stream.

3.1.2 baPuff + ba vs. baPuff + pa

The second prediction was that infants would look longer to trials where the baPuff + ba stimulus stream was presented than to those in which they were presented with baPuff + pa. The assumption behind this predicton is similar to that outlined above: the co-presentation of an puff of air with the unaspirated /ba/ token would render that token more like an aspirated /pa/ token. Thus, when the baPuff token was paired with the unaspirated token the stream would seem more alternating. In contrast, when the baPuff token was presented with the aspirated token, the stream would be perceived as less alternating. As Figure 4 illustrates, the infants behaved as predicted in that they looked significantly longer to the baPuff + ba stimulus stream. Again, if the infants were only paying attention to the acoustic signal and not integrating the aerotactile cues, then they would have perceived the baPuff + pa stimulus stream as the more alternating of the two and thus have looked longer to it. That they didn't indicates that they were integrating the multimodal cues.

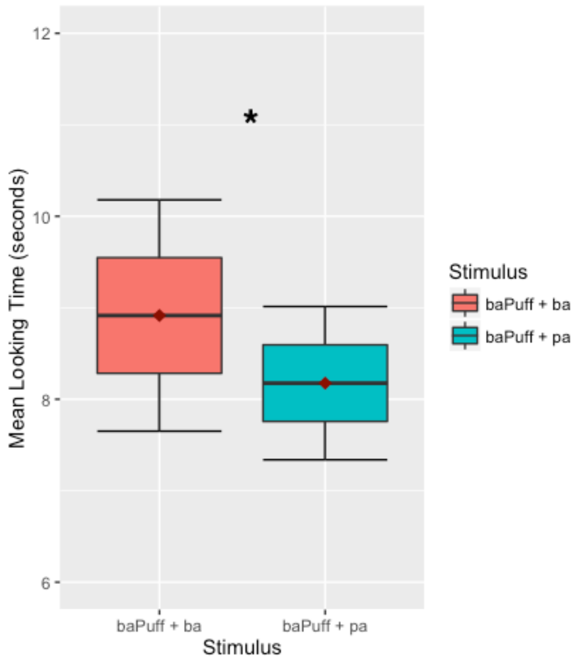

Figure 4: This boxplot compares mean looking times for the two stimulus streams baPuff + ba and baPuff + pa. As can be seen in the figure, the infants looked significantly longer to the baPuff + ba stimulus stream.

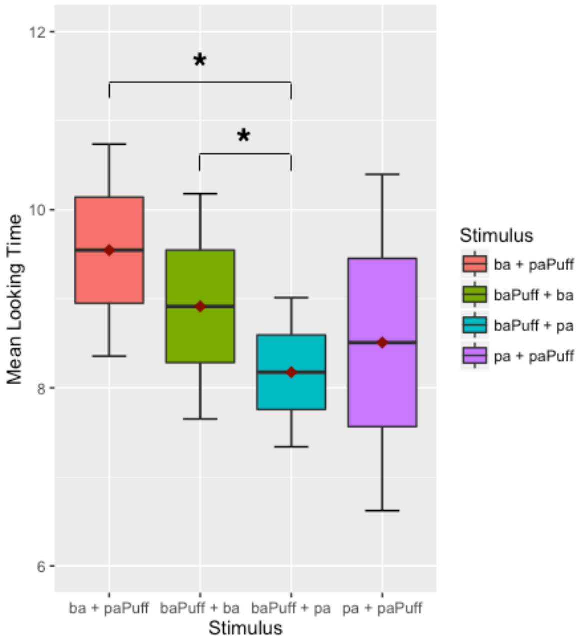

Figure 5: Boxplot comparing the mean looking times for all four stimulus streams. This figure clearly shows the increased variability in the infant response to the pa + paPuff stimulus stream as compared to the other stimulus streams.

No other factors emerged as significant predictors of looking time; the infants did not show significant differences between the two alternating stimulus streams (baPuff + ba and ba + paPuff) or the two non-alternating streams (baPuff + pa and pa + paPuff). In addition, there was no significant difference between either of the alternating streams and the non-alternating stream pa + paPuff, however. It is worth noting that there was a great deal of variability in how the infants responded to the pa +paPuff trials (see Figure 5), much more so than for any of the other stimuli. Several factors may be at play here, and I will return to this result in the discussion below.

4 Discussion

The current study tested the hypothesis that speech production experience is necessary to be able to integrate auditory and aero-tactile cues during speech perception. The current findings do not support this hypothesis. Recall that the overarching prediction was that, if the infants were integrating the multisensory cues, the co-presentation of synchronous airflow with an unaspirated /ba/ token would make the infants perceive the token as more like an aspirated /pa/. Further, this would influence how alternating the infants perceived a given stimulus stream to be and thus affect how long they looked during the trial. As discussed in the previous section, the results show that the infants looked less when a /ba/ accompanied by airflow was paired with a plain /pa/ (baPuff + pa) than when it was paired with a plain /ba/ (baPuff + ba) in a stimuli stream. Critically, this pattern of looking supports only the predictions of a hypothesis in which the preverbal infants were integrating the auditory and aero-tactile cues. As mentioned above in the predictions, if infants had been attending to only the auditory signal, we would have predicted the opposite looking patterns for these two stimulus streams. Similarly, when the auditory signal was held constant and the location of the puff was manipulated (i.e., baPuff + pa vs. ba +paPuff), the infants' perception of the streams was affected: instead of treating the two stimulus streams equivalently (as predicted if they do not integrate), the infants looked significantly longer when the aerotactile cue occurred on the aspirated token. Finally, as noted in the results section, there was no significant difference in looking times to the two alternating stimulus streams or between the two non-alternating streams. Together these results show that the preverbal English-acquiring infants tested were influenced by airflow cues much like the adults in the original Gick and Derrick study; when they were presented with unaspirated /ba/ tokens accompanied by a puff, they perceived them to be more like an aspirated /pa/. These findings suggest that perceivers do not require experience feeling their own airflow during productions of aspirated and unaspirated tokens in order to integrate auditory and aerotactile information. Moreover, these results are in keeping with the evidence outline in the introduction that infants integrate other multi-sensory speech cues well before the begin speaking (e.g., audiovisual speech). The current study offers additional evidence that infants have some (likely unconscious) knowledge of the sensory output of articulator movements before they begin babbling, contrary to what has been proposed in some computational models of speech production (for example, see Guenther, 1994; Guenther, 1995; Tourville & Guenther, 2011).

As mentioned in the results above, the infants did not appear to treat the /pa/ accompanied by a puff of air as equivalent to a /pa/ without. Though there was no statistical difference between the two "non-alternating" streams (i.e., pa + paPuff and baPuff + pa), there was considerable variability both within and across infants as to how they responded during pa + paPuff trials. A few points warrant discussion with respect to this finding. First, we have no a priori reason to expect infants to behave in the same manner as adults in previous studies, especially given that the tasks differed across the two populations. It is important to keep in mind that the infants performed a discrimination task, which is markedly different from the two-alternative forced choice task the adult participants Gick and Derrick (2009) took part in. We do not know how adults would treat this comparison in a task that does not force them to assign a category label to what they heard.

Second, recall that the prediction for the paPuff token was not as strong as for baPuff. The alternate possibility noted in the introduction was that, instead of treating the airflow as a redundant cue to the aspiration in the acoustic signal, the infants might treat the paPuff token as a separate phonetic category. It is possible that the puff is having an additive affect to the aspiration cue present in the acoustic signal and that this is creating a strongly aspirated /pa/. Cross-linguistic evidence supports this possibility. For example, Korean has a three-way contrast for voice onset time: plain, aspirated, and tense. Of interest to the current findings, this contrast is not one of pre-voiced, voiceless unaspirated, and voiceless aspirated. Instead, Korean shows a contrast of voiceless unaspirated[2], voiceless slightly aspirated, and voiceless strongly aspirated stops (Cho, Jun, & Ladefoged, 2002). This three-way contrast strongly mirrors the way the stimuli from the current experiment contrast if indeed the airflow created a super-aspirated /pa/. Moreover, the infants in this study are still undergoing perceptual narrowing and may not have a developed their native phonetic categories. Results from a set of experiments by Burns, Yoshida, and Werker (2007) offer evidence that 6- to 8-month olds may not yet show language-specific VOT perception. Thus, the infants in the current study may have shown more variable reactions to this stream because they were less sure whether the pa + paPuff streams contained two separate phonetic categories.

A second, and related, possibility is that the presence of aerotactile cue pushed the infant's perception of the plain /pa/ toward the unaspirated stop /ba/ when directly compared. Perhaps the airflow cue was so salient that it overrode the cue in the acoustic signal on the plain token. While interesting, however, the question of whether the airflow made the paPuff token seem super-aspirated or the plain token seem unaspirated does not bear on the current research question. The fact that the addition of airflow made the same auditory token seem sufficiently different to the infants indicates that they were integrating the cross-modal cues.

Though the results of the current study take us a step closer to understanding the origins of audio-aerotactile integration, we are left with an obvious question: if not through production experience, how does the ability to integrate auditory and aerotactile information arise? While the current study was not designed to tease apart any remaining options, it is useful to consider what some of those options might be and how future work could investigate their role in audio-aerotactile integration. One possibility worth considering is the infant's experience as a language perceiver. As mentioned in the introduction, preverbal infants like those in this study do not have experience feeling their own aspiration during the production of aspirated and unaspirated stops. However, because they are growing up in an English environment, they likely have experience feeling their caregivers' speech while listening to these sounds. Caregivers often hold their infants quite closely, and infants could be learning through their caregivers' airflow. Through close contact with caregivers who are speaking, they may discover that some sounds are accompanied by sudden bursts of airflow while others are not. Unfortunately, the dearth of data regarding exactly how closely caregivers speak to their infants—and how often—makes it difficult to do

[2] This is true when the stops are in syllable position like those in the stimuli for the current study. It should be noted, however, that these stops are realized as voiced inter-vocalically.

more than speculate. However, future work could investigate the potential influence of perceptual experience by testing preverbal infants growing up in a language environment with aspiration.

Of course, it may not be the case that the ability to integrate audio and aerotactile information emerges through a single mechanism or experience. The infants may be learning about airflow and sound through a variety of different experiences in both perception and production. For example, though infants at this age may just be beginning the babbling phase, they have certainly been vocalizing and exploring different articulations for several months. While this very limited production experience may not provide direct information about aspirated and unaspirated stops, it could be one method through which they discover that airflow and the sounds coming out of their mouths are related. They may also have access to the visual effects of airflow on environment. Previous work has shown that visual representations of aspiration (e.g., the flickering of a candle) can influence adult native English speakers in much the same way as direct aero-tactile sensations (Mayer, Gick, Weigel, & Whalen, 2013). While infants are unlikely to have conscious awareness of this sort of environmental effect of aspiration, it may nonetheless be included in their general knowledge of aero-tactile speech information.

While the results from the current study are compelling, certain steps must be taken before drawing strong conclusions. To start, data collection is ongoing. This paper reports results from a relatively small sample of only twelve babies—a number half the eventual sample. Second, a control group may need to be run to ensure that infants are responding to the tactile sensation of the airflow and not the sound of the puff exiting the tube. Finally, the reliability of the current coding needs to be verified. To this end, a second coder will independently code a random selection of 25% of the infant videos. Her coding results will then be compared with the coding used for the current analysis to assess agreement and ensure that the results reported in this paper indeed reflect the infants' behavior and are not a result of unconscious coder bias.

5 Conclusion

In sum, the results reported in this paper show that infants integrate audio and aerotactile cues well before they being producing aspirated and unaspirated stops. Moreover, this integration appears to arise before the infants have begun babbling, the phase in which it has been argued infants that infants learn the relationship between sounds, their articulations, and the corresponding sensory output. Though many questions remain regarding the origin and developmental trajectory of this multisensory integration, the current study constitutes an important step not only in demonstrating that infants integrate these cues, but also in developing a novel method to test this ability that can be applied to future work.

References

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child development*, *52*(4), 1135.

Audacity Team (2016). Audacity (Version 2.1.2) [Computer Program]. Retrieved January 12, 2016, from http://audacityteam.org

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. <doi:10.18637/jss.v067.i01>.

Best, C., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, *21*, 295.

Bicevskis, K., Derrick, D., & Gick, B. (2016). Visual-tactile integration in speech perception: Evidence for modality neutral speech primitives. *The Journal of the Acoustical Society of America*, *140*(5), 3531-3539.

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015*).* Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 112(44), 13531–13536. http://doi.org/10.1073/pnas.1508631112*

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*(03), 455-474.

Byun, T. M. (2012). Bidirectional perception–production relations in phonological development: evidence from positional neutralization. *Clinical linguistics & phonetics*, *26*(5), 397-413.

Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of phonetics*, *30*(2), 193-228.

Cohen J.D., MacWhinney B., Flatt M., and Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257-271

Derrick, D., O'Beirne, G. A., De Rybel, T., & Hay, J. (2014). Aero-tactile integration in fricatives: converting audio to air flow information for speech perception enhancement. *INTERSPEECH* (pp. 2580-2584).

Eimas, P. D., Siqueland, E. R., Juscyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*(7272), 502-504. doi:10.1038/nature08572

Gick, B., Jóhannsdóttir, K. M., Gibraiel, D., & Mühlbauer, J. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *The Journal of the Acoustical Society of America*, *123*(4), EL72-EL76.

Goldenberg, D., Tiede, M. K., & Whalen, D. H. (2015). Aero-tactile influence on speech perception of voicing continua. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, eds. The Scottish Consortium for ICPhS 2015 Glasgow, UK: the University of Glasgow.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological cybernetics*, *72*(1), 43-53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological review*, *102*(3), 594.

Hitchcock, E. R., & Koenig, L. L. (2013). The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research*, *56*(2), 441-457.

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, *106*(4), 1245-1248.

Kuhl, P. K., & Meltzoff, A. N. (1982, December). The bimodal perception of speech in infancy. American Association for the Advancement of Science.

Mayer, C., Gick, B., Weigel, T., and & Whalen, D. (2013). "Perceptual integration of visual evidence of the airstream from aspirated stops." Can. Acoust. 41(3), 23-27.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. doi:10.1038/264746a0

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237-247.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6, 191–196.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Attention, Perception, & Psychophysics*, *59*(3), 347-357.

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. European Journal of Neuroscience, 33: 2329–2337. doi:10.1111/j.1460-9568.2011.07685.x

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and cognitive processes*, *26*(7), 952-981.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234-43. doi:10.1016/j.cognition.2009.08.010