2019

Interactions Between Contextual and Phonetic Information in Children's and Adults' Perception of Non-Native Speech

Alexis P. Zosel
University of Mississippi

INTERACTIONS BETWEEN CONTEXTUAL AND PHONETIC INFORMATION IN CHILDREN'S AND ADULTS' PERCEPTION OF NON-NATIVE SPEECH

by
Alexis P. Zosel

A thesis submitted to the faculty of the University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2019

Approved by

____
Advisor: Professor Susan Loveall

____
Reader: Professor Kara Hawthorne

____
Reader: Professor Carolyn Higdon

ABSTRACT

ALEXIS PAIGE ZOSEL: Interactions Between Contextual and Phonetic Information in Children's and Adults' Perception of Non-Native Speech
(Under the direction of Susan Loveall and Kara Hawthorne)

Phoneme categorization (i.e., the ability to differentiate between different speech sounds) is not an easy task, as individuals must integrate multiple sources of information, including both acoustic and contextual information. When a talker has a foreign accent, the listener may place more weight on sentential context because of the ambiguity of the acoustic/phonetic information. For example, listeners are more likely to classify a word that is phonetically ambiguous between *goat* and *coat* as *goat* if the sentential context is "The boy milked the ___," and this is especially the case when listening to a non-native (compared to native) talker (Schertz & Hawthorne, 2018). In the present study, we extended this result in adults using multiple talkers and two tasks. Additionally, we continued this research with children and found that they do use sentential context, but, unlike adults, do not show significant differences in how they use context for native vs. non-native talkers. The findings of this study allow for a better understanding of how listeners of different ages flexibly reweight acoustic and contextual cues within the context of different types of talkers. Understanding this could allow for better communication in a diverse America.

Keywords: speech perception, non-native speech, children, cue weighting, phoneme categorization

TABLE OF CONTENTS

LIST OF FIGURES

I. Introduction

Phoneme categorization is the ability to classify different phonemes (individual units of sound which distinguish word meaning, e.g., /k/ vs /g/). It is one of the most fundamental tasks involved in speech perception; however, categorizing phonemes is not always an easy task. In everyday speech, sounds often run together, and sounds that are classified as the same phoneme may be acoustically distinct. For example, the /k/ in *skit* is acoustically more similar to /g/ than the /k/ in *kit*. Listeners must integrate many sources of information, such as acoustic/phonetic information, other sensory cues, and contextual information (e.g., Holt & Lotto, 2010; Yost, 1992). However, it's not clear how listeners weight each of these cues when they conflict or how this changes when listening to a non-native talker, whose speech may have less reliable phonetic cues. The purpose of this study is to determine how children and adults categorize phonemes by examining their weighting of different sources of information across multiple accents. Understanding the weight each cue is given when listening to a native versus non-native talker across age could offer important insight into better communication strategies amongst an every-increasingly diverse America.

A. Integrating Cues for Phoneme Categorization: Adults

Adult listeners integrate many sources of information (e.g., acoustic, visual, tactile, contextual) when determining what phoneme they hear. These sources are typically sensory, expectation-driven, or a combination of both. The primary sensory cue for speech perception is the acoustic signal, but other sensory cues (i.e., visual and tactile) can also influence phoneme categorization. In addition, the way listeners categorize a phoneme is influenced by their expectations (i.e., context-driven, talker-driven), which involves utilizing their previous knowledge to make predictions about what they are hearing. The following sections review each of these cues in detail. Because the current study tests categorization of voiced and voiceless stops (i.e., /k/ vs /g), the following discussion of cues for phoneme categorization largely focuses on cues that are relevant for stop voicing.

A.1 Acoustic Cues

The acoustic signal is the most obvious source of information in phoneme categorization. Perhaps the most thoroughly investigated acoustic cue to phoneme categorization for voiced vs. voiceless stop consonants is voice onset time (VOT). VOT is a continuous variable, measured as the time the vocal folds take to begin vibrating after releasing a stop consonant. Voiced stop consonants have lower VOT (<35ms), while voiceless stop consonants have higher VOT (>35ms). In other words, the vocal folds start vibrating more quickly after the burst release for a voiced stop versus a voiceless stop. VOT boundaries vary cross-linguistically, as the boundary between voiced and voiceless consonants is lower in Spanish than English even though both languages utilize a range of VOT values (Flege & Eefting, 1987).

While VOT is perhaps the most obvious cue to stop voicing, since it relates to vocal fold vibration onset, several other cues also influence how we distinguish voiced from voiceless stops. Aspiration (expelling a small puff of air upon burst release) is another acoustic cue for stop voicing that is closely related to VOT in English. Higher VOT leads to more aspiration and the perception of voicelessness, while lower VOT leads to less aspiration and the perception of voicing (Deterding & Nolan, 2007).

Perhaps more surprisingly, pitch and duration are also acoustic cues to stop voicing. Haggard, Ambler, and Callow (1970) found that as pitch increases during the onset of voicing, the listener may be more likely to categorize the preceding stop as voiceless, and a low pitch during onset of voicing may cause the listener to categorize the preceding stop as voiced. Specifically, Whalen, Abramson, Lisker, & Mody (1993) found this to be the case even when VOT is held constant. In addition, Raphael (1972) revealed that the shorter the vowel preceding a stop consonant, it was more likely to be perceived as voiceless, and vice-versa. Therefore, a listener integrates multiple acoustic cues – not all of which are related to vocal fold vibration – when determining whether they are listening to a voiced or voiceless stop.

A.2 Other Sensory Cues

While acoustic cues play the dominant role in phoneme categorization, other sensory cues are also relevant for determining stop voicing. Gick and Derrick (2009) found that the illusion of aspiration through tactile cues affects the ways in which adult listeners perceive stop voicing. Listeners were more likely to categorize a /b/ as a /p/ when they felt a small puff of air on their hand or neck because voiceless stops (i.e., /p/) produce aspiration, while voiced stops (i.e., /b/) do not.

Still, other sensory cues contribute to the way in which we categorize phonemes. The well-known McGurk effect, which shows how our perception of the sounds we hear is shaped by the visual input, is a prime example of multimodal sensory cues being integrated (McGurk & MacDonald, 1976). For example, if a listener hears /ba/ but sees the mouth movements for /va/, the visual input will override the acoustic input, and they will perceive it as /va/. This remains the case even when the listener is aware that this effect is taking place. Though we are not testing the impact of visual or sensory cues in the current study, these examples are pertinent because they exhibit how we integrate various types of cues when perceiving speech.

A.3 Context-Driven Expectations

In addition to sensory information, expectation-driven information influences categorization of phonemes. Contextual cues are one important source of expectation-driven information that become increasingly important when the listener is presented with unclear or missing information (Borsky, Tuller, & Shapiro, 1998; Connine &

Clifton, 1987; Davis & Johnsrude, 2007; McClelland, Mirman, & Holt, 2006). In their experiment, Borsky and colleagues (1998) tested 33 monolingual, English-speaking undergraduates who listened to stimuli with a sentential context bias toward either *goat* or *coat*. Each sentence ended in *_oat* with the first phoneme varying along a 10-step VOT continuum from 10ms (a good initial /g/) to 64ms (a good initial /k/). Participants heard these sentences on a computer and decided whether what they saw on the screen afterward (either the visual probe GOAT or COAT) matched or did not match what they heard at the end of each sentence. Borsky et al. found that young adults relied on the sentential context over the acoustic (VOT) details when acoustics were ambiguous (i.e., in the 28-34ms range).

A.4 Talker-Driven Expectations

Another expectation-driven cue that influences phoneme categorization is the listener's expectations about the talker. For example, to test how adult listeners categorize the fricatives /s/ and /ʃ/, participants in Strand (1999) first listened to auditory-only stimuli and then audiovisual stimuli (where they saw faces speaking these stimuli). She found that gender stereotypes affected the way listeners perceived the phonemes. A female face accompanying the auditory input caused the boundary between /s/ and /ʃ/ to shift up in frequency, whereas a male face accompanying the auditory input caused the boundary between the two phonemes to shift down in frequency.

Beyond gender, accent also influences how a listener categorizes an acoustic signal. For instance, non-native speech is sometimes less intelligible and comprehensible than native speech, as increased errors may occur in prosody, phonetics, phonemics, and/or grammar (Munro & Derwing, 1995). Recent evidence from Schertz and Hawthorne (2018) suggests then that listeners adjust their phoneme categorization strategies and their weighting of cues when listening to a non-native talker. As this study investigated the integration of multiple speech perception cues, it is discussed further in the following section.

A.5 Integrating Acoustic, Contextual, and Talker-Specific Information

Borsky et al. (1998) found that adult listeners weighted the contextual information higher than the acoustic information when acoustics were ambiguous due to having a VOT near the perceptual boundary between voiced /g/ and voiceless /k/. However, they only examined native speech, not non-native speech. Since non-native speech is often less intelligible in other ways, listeners may further shift their weighting of cues by paying more attention to the sentential context.

To test this hypothesis, Schertz and Hawthorne (2018) conducted a study utilizing a 9-step VOT continuum from 5ms (a good initial /g/) to 85ms (a good initial /k/) to test how participants weighted acoustic cues and sentential context when listening to a non-native versus native talker. Like Borsky et al., (1998), Schertz and Hawthorne (2018) used sentences biased to end in either *goat* or *coat*, such as “The wise grandmother remembered to wear the _____” or “The handsome man forgot to feed the _____.”

Listeners (all native English speakers) heard recordings of both a native English talker and a native Mandarin Chinese talker saying the sentences and then saw either the visual probe GOAT or COAT on the screen. Listeners then indicated if the word they saw on the screen matched or did not match the last word in the sentence to which they just listened.

While Schertz and Hawthorne found that acoustic cues were the most important in categorizing the phonemes overall, they also found listeners weighted the sentential context more highly for the non-native talker. This suggests that when speech becomes less clear in other ways (i.e., ambiguous VOT, lower intelligibility), listeners may depend more on higher-level information, such as sentential context. However, with increased exposure to the non-native talker, individuals were able to utilize adaptation strategies to place less reliance on the context and more on the phonemic characteristics of that accent (Schertz & Hawthorne, 2018).

B. Integrating Cues for Phoneme Categorization: Children

Children's integration of speech perception cues (and, specifically, the ways in which they process non-native speech) is less explored than adults'. This is problematic because the number of people speaking a language other than English in the home is increasing rapidly (Ryan, 2013). Children generally suffer more in speech comprehension in adverse listening conditions than adults (e.g., Fallon, Trehub, & Schneider, 2000, 2002). Specifically, research has found children have increased difficulty in processing non-native accents (Bent, 2014; Bent & Atagi, 2015; Holtby, 2010; McDonald, Gross, Buac, Batko, & Kaushanskaya, 2017). Barker and Turner (2015) found that task played a role in children's perception of non-native speech, with children more easily recognizing words spoken by a native talker, yet better comprehending a story narrated by a non-native talker. Increased exposure and context that come from narrating a story paired with a novel accent, which might have cued the children in to pay more attention, may be the reason for the interesting result of children better comprehending a story narrated by a non-native talker.

Like adults, children utilize multiple sources of information when categorizing phonemes. Studies show that infants as young as one to four months have the ability to discriminate between voiced and voiceless stop phonemes based on their VOT values (Clarkson, Eimas, & Marean, 1989; Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Additionally, children have been shown to incorporate other sensory cues when categorizing phonemes. For example, Rosenblum, Schmuckler, and Johnson (1997) tested the McGurk effect on infants of five months and found that the visual input interfered with perception of the acoustic input, just as has been seen in adults (McGurk & MacDonald, 1976).

Children use sentential context in phoneme categorization as well. Holt and Bent (2017) examined 168 children listening to sentences with high-predictability vs. low-predictability contexts spoken by native and non-native talkers. A high-predictability context is when the sentence has a clear bias toward a certain word (e.g., "The boy milked the ___"), while a low-predictability context is when the sentence does not have a clear bias toward a certain word (e.g., "The girl wanted to ___"). They found that children ages 5-7 years were better at repeating native speech (vs. non-native speech), suggesting that native speech is easier for children to process. High-predictability contexts produced better scores than low-predictability contexts for both native and non-native speech, and the older the child, the greater the benefit of sentential context. Yet, they did not find significant differences between the children's benefit from context for the native vs. non-native speech. While this indicates that children are integrating contextual information, it does not provide evidence that they are making use of talker-specific information about the accent, since whether they heard a native or non-native talker did not impact their ability to repeat the speech (Holt & Bent, 2017).

Though less is known about the ways in which children integrate multiple sources of information, some research has been done in this domain. Creel, Rojo, and Paullada (2016) explored preschool-aged children's integration of contextual and visual cues in comprehending non-native speech. Over the course of four experiments, the children listened to native and non-native talkers say sensical and non-sensical sentences (based on their context) and then were asked to choose which of four visuals matched what they heard (Experiments 1 and 2) or repeat words (Experiments 3 and 4). While children in Experiment 1 had both sentential context and visual context available to them, sentential context was taken away for children in Experiment 2, visual context was taken away for children in Experiment 3, and both sentential and visual context were taken away for children in Experiment 4. Creel, Rojo, and Paullada (2016) found that children struggled more with comprehension of non-native speech (compared to native speech) when support from *only* visual aids was removed or sentential context *and* visual aids were removed. In summary, Creel and colleagues found effects of talker (native vs. non-native) and visual context on children's ability to comprehend the speech, but no effect of sentential context. In their review article, Cristia et al. (2012) note that research within this domain is limited and more needs to be done.

C. The Current Study

It has been well-documented that multiple cues are integrated by adult listeners when deciding what phoneme they hear; in particular, acoustic cues, contextual cues, and the talker's accent can all play a role. There is initial evidence that this is also the case for children. In this study, we extended previous work by Schertz and Hawthorne (2018) by using a different task with multiple talkers and by taking a developmental perspective.

The purpose of this paper is to determine how listeners integrate multiple cues in phoneme categorization and how this changes developmentally. We have two research questions: 1) How do adults integrate acoustic (VOT) and contextual cues when listening to native vs. non-native talkers? 2) Do children ages 5-13 years show adult-like weighting of acoustic and contextual cues when listening to native and non-native talkers?

In this study, we tested these research questions using a phoneme categorization task. Participants listened to sentences spoken by native and non-native talkers with a sentential context bias to end in *goat* or *coat*, e.g., "The wise grandmother remembered to wear the _____" or "The old grandfather liked to milk the _____." Each sentence ended in _*oat*, with the first phoneme varying along a VOT continuum from a good initial /g/ to a good initial /k/. Listeners then indicated whether they heard *goat* or *coat* by pressing a key on the keyboard, pointing, or verbally responding.

Because adult listeners may expect the acoustics of non-native speech to be less reliable, we hypothesized that adults would rely on sentential context more when listening to non-native talkers vs. native talkers. Such a result would replicate the findings of Schertz and Hawthorne (2018) and extend them to multiple talkers of each accent. On one hand, children struggle with comprehension of non-native speech in certain task designs, so we may find that they rely even more on sentential context than adults when listening to non-native talkers. On the other hand, based on previous work indicating children exhibit similar reliance on sentential context for native and non-native speech (i.e., Holt & Bent, 2017), we may find children show no differences in their weighting of sentential context across native and non-native speech.

II. Experiment 1: Picture Selection

A. Methods

A.1 Participants

We recruited 24 undergraduates (one male; ages 19-22 years) from the University of Mississippi to participate in our experiment. Participants were monolingual English speakers with normal vision and hearing and no significant language impairments. This information was self-reported on a language background questionnaire (Appendix A). The questionnaire consisted of 12 questions, including questions about exposure to other languages. This information was used to identify participants who did not meet inclusion criteria, as well as those who may have had significant exposure to Mandarin Chinese, the non-native accent that was used in this study. The language background questionnaire was given to participants after completion of the experiment to ensure participants were not alerted to the fact that they'd be listening to a Mandarin-accented talker in the experiment. Most participants received extra credit in exchange for their participation, though a few volunteered their time.

A.2 Procedures

The study was approved by the IRB at the University of Mississippi and took place at a lab on campus. Before the experiment, the participant signed an informed consent form.

Participants completed a phoneme categorization task, adapted from Schertz and Hawthorne (2018), which, in turn, was adapted from Borsky et al. (1998). For each trial, the participant heard a sentence ending in a VOT continuum between *goat* and *coat*. After the sentence, pictures of a coat and a goat appeared on the screen. Participants then chose which picture matched what they heard by pressing a key on a laptop keyboard. (For practice trials participants sometimes responded by pointing or responding verbally.) Note that both Schertz and Hawthorne (2018) and Borsky et al. (1998) used a match/ mismatch task. That is, participants saw either the visual probe word GOAT or COAT after listening to each sentence and decided whether the word matched or did not match the last word they heard. We changed to a picture selection task because we thought it might require less cognitive work for the child participants. In total, the session took around 45-50 minutes.

A.3 Materials

A.3.1 Auditory Stimuli

Sentences were recorded by six talkers using a Countryman E6 microphone and a Zoom H4N recorder. Three talkers were native English speakers, and three were native Mandarin Chinese speakers. Native Mandarin talkers were students in Canada and had lived there for nine, two, and two years at time of recording. Native English talkers were students born in Ontario, Canada. All participants were living in Toronto at time of recording.

Each talker produced four sentences that could be completed by *coat* (but not *goat*) and vice-versa, for a total of eight unique sentences (Table 1). An example of a *coat*-biased sentence would be "The young girl learned how to put on the _______," while an example of a *goat*-biased sentence would be "The young girl learned how to milk the _______."

***Coat*-Biased Sentences**	***Goat*-Biased Sentences**
The wise grandmother remembered to wear the c/goat.	The wise grandmother remembered to feed the g/coat.
The lazy brother didn't unbutton the c/goat.	The old grandfather liked to milk the g/coat.
The young girl learned how to put on the c/goat.	The young girl learned how to milk the g/coat.
The handsome man forgot to wear the c/goat.	The handsome man forgot to feed the g/coat.

Table 1. Experiment sentences.

One token of *the coat* was spliced from the end of a single sentence produced by each talker. This token was used to create an 8-step VOT continuum from the word-initial stop /k/ (*the coat*) to /g/ (*the goat*). The VOT values ranged from 0ms (a good /g/) to 70ms (a good /k/) in 10ms increments (e.g., 0ms, 10ms, 20ms, etc). Each talker's manipulated words were then inserted into the end of each of the eight sentences produced by that talker. There were a total of 384 sentences: 6 talkers (3 native, 3 non-native) * 8 sentences (4 *goat*-biased, 4 *coat*-biased) * 8 VOT steps. In addition, there were four practice trials involving different contextually biased sentences that ended in *goat* or *coat*.

A.3.2 Visual Stimuli

While participants listened to each sentence, they simultaneously saw a red dot on the screen. They then saw pictures of *both* a goat and a coat pop up on the screen (Figure 1) and had to select which picture matched the last word of the sentence they just heard. The pictures of the goat and coat were randomized each time with regard to which image appeared on the right versus the left.

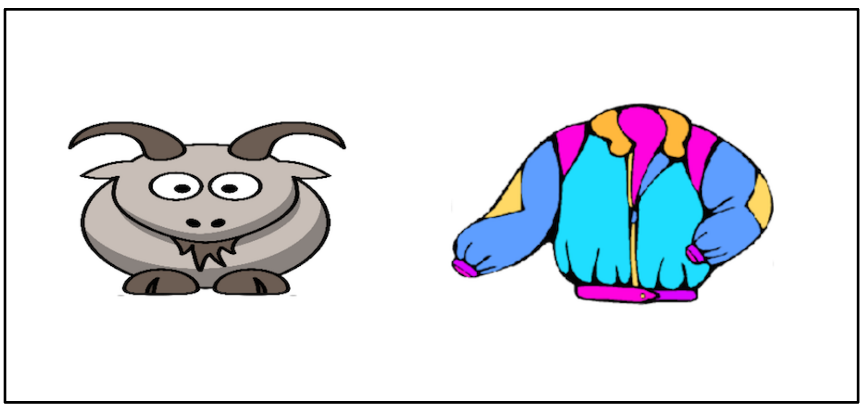

Figure 1. Visual display for practice and experimental trials.

A.3.3 Equipment

The experiment was implemented using PsychoPy, version 1.85.4. Sentences were presented via a Bose SoundLink Mini Bluetooth speaker II, model 416912.

A.4 Design

Participants were tested on one of four randomized lists, to help control for order effects**.** Each list was blocked by talker. Two of the lists began with a native English talker block, and the other two began with a non-native English talker block. Talker blocks then alternated from native to non-native talkers, and items were randomized within each block. Additionally, the location of each picture on the screen was randomized by trial.

The independent variables were talker (three native, three non-native), sentential context (four *goat*-biased, four *coat*-biased), and VOT (0-70ms in 10ms steps). The dependent variable was whether participants responded indicating *coat* or *goat*, translating to a /k/ or /g/ categorization of the initial phoneme.

B. Results

Responses were converted to /k/ (*coat*) or /g/ (*goat*) categorizations of the initial phoneme, based on whether participants indicated a *coat* or *goat* response. Responses were considered outliers and discarded if more than two seconds elapsed between the end of the sentence and the participant's response. This response time was chosen by examining a histogram of all response times. A total of 301 trials were cut (3%), leaving 9278 total trials.

Results are presented in Figure 2. The y-axis represents the percentage of *coat* responses for each type of talker based on VOT and sentential context. For example, 75% *coat* responses means participants were giving a /k/ response 75% of the time and a /g/

response 25% of the time. The lines in each graph represent whether the sentence was contextually biased to end in *goat* or *coat*. The gap between the coat-biased and goat-biased lines indicates the effect of context.

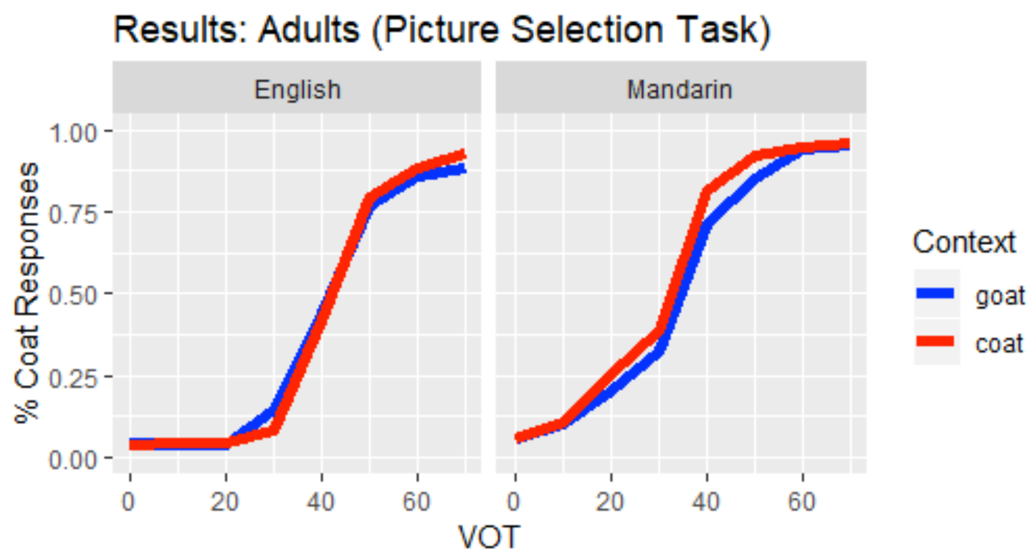

Figure 2. Responses across the VOT range for each type of talker for Experiment 1.

Responses were analyzed using logistic mixed-effects regression models in *R* (R Core Team, 2018) using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015). The model included fixed effects of talker L1 (native vs. non-native; contrast coded), sentential context (*coat*- vs. *goat*-biased; contrast-coded), and VOT (0-70ms in 10ms intervals; centered). The model also tested for an interaction between talker L1 and context. These fixed effects represent the independent variables of interest. In addition, the model included random intercepts for participant and item and random slopes for VOT and talker L1 by participant. Random effects by participant account for individual variability, while random effects by item account for item-level variability.

As expected, there was a significant main effect of VOT ($\beta = .15$, $SE = .013$, $p < .001$, odds ratio = 1.14), with more /k/ (*coat*) responses for higher VOT values, and a significant main effect of context ($\beta = .24$, $SE = .084$, $p = .0043$, odds ratio = 1.24), with more /k/ (*coat*) responses for *coat*-biased sentences. Additionally, there was a significant main effect of talker L1 ($\beta = 1.59$, $SE = .20$, $p < .001$, odds ratio = 4.70), with more /k/ (*coat*) responses for the Mandarin talkers. There was also a significant interaction of talker L1 by context ($\beta = .43$, $SE = .14$, $p = .0028$), indicating that the effect of context depended on whether the talker had a native or non-native accent.

Since there was a significant interaction of context by talker L1, simple effects of context were examined separately for each type of talker using the *testInteractions* function from the *phia* package (De Rosario-Martinez, 2015) in *R*. There was a significant simple effect of context for the Mandarin talkers ($\chi^2(1) = 17.02$, $p < .001$, odds ratio = 2.43), but no simple effect of context for the English talkers ($\chi^2(1) = .051$, $p = .82$, odds ratio = 1.30). This is in line with the prediction that adults would use sentential context more for the non-native talker; however, it differs from previous findings (i.e., Schertz & Hawthorne, 2018) in that context was not used regardless of L1 status.

C. Interim Discussion

Results partially replicate Schertz and Hawthorne (2018), in that adults relied more on the sentential context when listening to the non-native talkers. However, unlike Schertz and Hawthorne (2018), we did not find a significant effect of context for the native English talkers. Our failure to replicate this effect could be due to differences in the stimuli. Our experiment used different (and multiple) talkers for the native and non-native speech, and our VOT continuum increments and number of steps were slightly different. This could impact results, as participants might be responding according to a specific talker or specifics within the VOT continuum. On the other hand, it could be because of the change in task. While Schertz and Hawthorne (2018) and Borsky et al. (1998) both used a match/mismatch task, Experiment 1 used a more child-friendly picture selection task.

In order to determine whether the lack of context effect of the native talkers was due to our use of a picture selection task, a match/mismatch task was used for Experiment 2. However, unlike Borsky et al. (1998) and Schertz and Hawthorne (2018), participants matched a photo rather than a probe word to maintain child-friendliness (Experiment 2.II).

III. Experiment 2: Match/Mismatch

2.I. Adults

A. Methods

A.1 Participants

Participants were 26 undergraduates (all female, ages 19-22 with the exception of one 31-year-old). See Section II.A.1 for additional information about recruitment and inclusion criteria. All participants completed the language background questionnaire mentioned in Section II.A.1 and received extra credit in exchange for their participation.

A.2 Procedures, Equipment, Materials, and Design

The same basic procedures, stimuli, equipment, and design of Experiment 1 were used for Experiment 2.I, except for the task change. Unlike Experiment 1, in which participants selected a picture of a coat or a goat to indicate whether each sentence ended in *goat* or *coat*, participants in Experiment 2.I indicated whether a single picture matched or did not match the last word in the sentence they just heard. In addition to the probe picture (the goat or coat in the middle of the screen), there was also a green checkmark (bottom left corner of the screen) and red *X* (bottom right corner of the screen). Participants selected the checkmark if the picture matched what they heard or the *X* if the picture mismatched what they heard. The checkmark and *X* stayed in the same positions for each trial, but whether the goat or coat appeared on the screen was randomized by trial. As in Experiment 1, participants indicated their response by pressing a key on the keyboard (or by pointing or responding verbally for practice trials).

B. Results

Responses were converted to *coat* (/k) or *goat* (/g/) responses to indicate a categorization of the initial phoneme, based on whether participants indicated *same* or *different*. For example, when a participant saw a picture of a coat after hearing a sentence, a *same* response was converted to a *coat* (/k/) response, while a *different* response was converted to a *goat* (/g/) response. Outliers were trimmed using procedures described in Section II.B. A total of 295 trials were cut (3%), leaving 9497 total trials.

Results are presented in Figure 3.

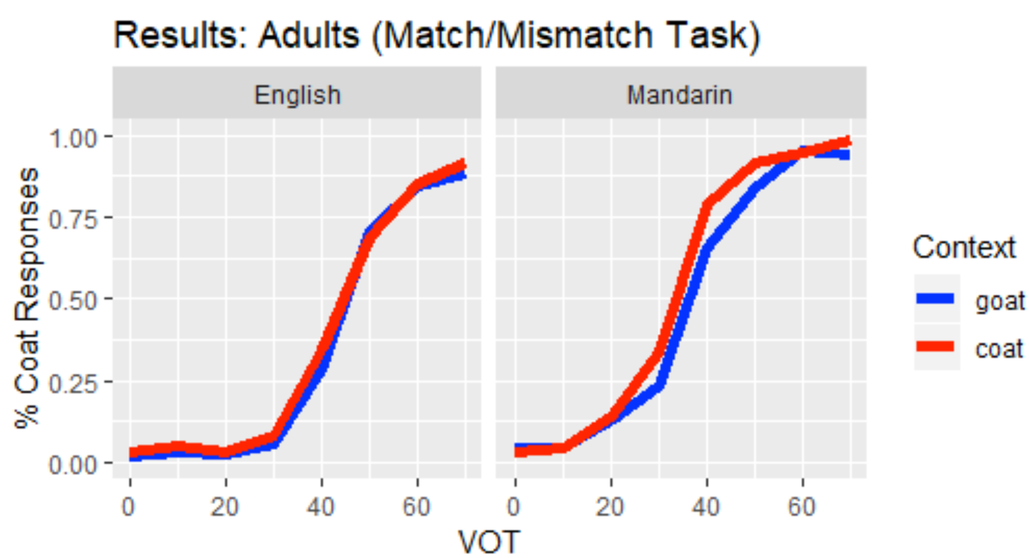

Figure 3. Responses across the VOT range for each type of talker for Experiment 2.I.

Data were modeled as described in Section II.B. Again, there was a significant main effect of VOT ($\beta = .13$, SE = .0073, $p < .001$, odds ratio = 1.13), with more *coat* (/k/) responses for higher VOT values, and a significant main effect of context ($\beta = .32$, SE = .067, $p < .001$, odds ratio = 1.35), with more *coat* (/k/) responses for *coat*-biased sentences. Additionally, there was a significant main effect of talker L1 ($\beta = 1.50$, SE = .16, $p < .001$, odds ratio = 4.04), with more *coat* (/k/) responses for the Mandarin talkers. There was also a significant interaction of talker L1 by context ($\beta = .26$, SE = .13, $p = .047$), indicating that the effect of context depended on whether the talker had a native or non-native accent.

To further explore the interaction of context by talker L1, simple effects of context were examined separately for each type of talker. There was a significant simple effect of context for the Mandarin talkers ($\chi^2(1) = 24.16$, $p < .001$, odds ratio = 2.42) and for the English talkers ($\chi^2(1) = 4.028$, $p = .045$, odds ratio = 1.27). While simple effects of context were significant for both types of talkers, the presence of a significant context by talker L1 interaction and the larger effect size for the Mandarin talkers indicates that listeners use context more for the non-native talkers. This is in line with previous findings from Schertz and Hawthorne (2018) that sentential context was still used regardless of L1 status, but that context is used more when listening to a non-native talker.

2.II. Children

A. Methods

A.1 Participants

We recruited 18 children (5 males; ages 5;8-13;8; mean age of 10;2 with a standard deviation of 2;3) from schools and by word of mouth around the Oxford, Mississippi area. Child participants were monolingual (n = 16) or bilingual from birth (n = 2) English speakers with normal vision and hearing and no significant language impairments. (Note that while we had no bilingual from birth adult participants in the previous experiments, they would have met our predetermined inclusion criteria). This information was reported by a parent/legal guardian on a language background questionnaire (Appendix B). Note that a large number of participants (n = 15) were children of faculty at the University of Mississippi. Children were offered a small toy, and children's parents were offered a small amount of money in the form of a gift card if they brought their child to the university lab.

A.2 Procedures, Equipment, Materials, and Design

Procedures were the same as for Experiment 2.I, with a few exceptions. First, while adults were run exclusively in the university lab, children were run either in a quiet room at their school or home, or in the university lab. Before the experiment, each child's parent/legal guardian signed an informed consent form. Children six years of age or older also gave verbal assent. Second, child participants were offered the option of tracking their progress with stickers or crayons. Children were also allowed breaks as needed.

The most significant change was to the number of experimental trials, which was reduced. The number of talkers was reduced from 6 to 4 (two native and two non-native) and the number of sentences from 8 to 6 (three *goat*-biased and three *coat*-biased). As opposed to Experiment 1 and Experiment 2.I which totaled 384 sentences, Experiment 2.II consisted of 192 sentences: 4 talkers * 6 sentences * 8 VOT steps. Still, there were four practice trials involving different contextually biased sentences that ended in goat or coat.

Finally, while some children indicated their response by pressing keys on a keyboard like adult participants, others (i.e., younger children) responded verbally. Verbal responses were *same* or *different* to indicate whether the goat or coat they saw on the screen matched or did not match the last word in the sentence they just heard.

B. Results

Data were analyzed as described in Section III.(2.I).B. Again, outliers were trimmed using procedures described in Section II.B, except responses were considered outliers if more than four seconds elapsed between the end of the sentence and the child's response. A total of 293 trials were cut (9%), leaving 3110 total trials. Results are presented in Figure 4.

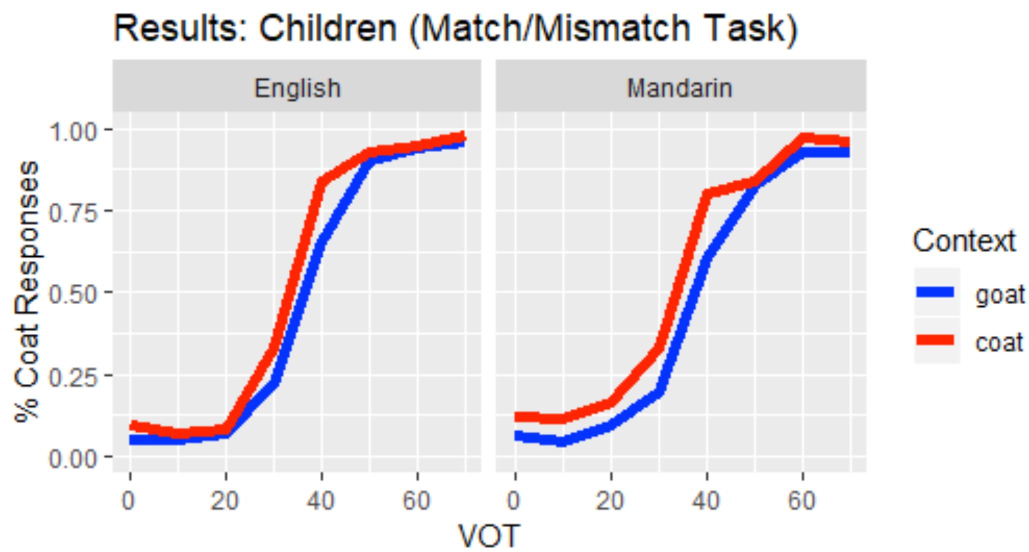

Figure 4. Responses across the VOT range for each type of talker for Experiment 2.II.

Data were modeled as described in Section II.B. As expected, there was a significant main effect of VOT ($\beta = .15$, SE = .016, $p < .001$, odds ratio = 1.15), with more *coat* (/k/) responses for higher VOT values. There was no significant interaction of talker L1 by context ($\beta = .23$, SE = .23, $p = .33$), indicating that the effect of context did not depend on whether the talker had a native or non-native accent. In fact, there was no significant main effect of talker L1 ($\beta = -.046$, SE = .22, $p = .83$). There was, however, a significant main effect of context ($\beta = .66$, SE = .18, $p < .001$, odds ratio = 1.90), with children responding *coat* (/k/) more for *coat*-biased sentences, regardless of talker.

IV. Discussion

Our first research question asked how adults integrate acoustic (VOT) and contextual cues when categorizing phonemes while listening to native vs. non-native talkers. Results from Experiment 2.I suggest that all three sources of information play a role in categorizing phonemes. Unsurprisingly, acoustics (VOT) play the most important role, as adults largely categorized the phoneme as a /g/ in the lowest VOTs and a /k/ in the highest VOTs across both talkers. We also see, though, that adults integrate sentential context into their phoneme categorizations. Across both talkers, adults were more likely to perceive the phoneme that (when paired with _*oat*) made sense in the sentential context. Finally, adults vary their weighting of sentential context based on the talker. Adults in Experiment 2.I used context more when listening to the non-native talkers than the native talkers, perhaps because they were expecting the non-native talkers to be less phonetically reliable. These results replicate the findings of Schertz and Hawthorne (2018) and extend them to multiple talkers of each accent. By replicating this effect with multiple talkers of both native and non-native accents, we can be more confident that the difference in context effect is a result of whether the accent is non-native compared to native, rather than some result of the individual talkers themselves.

In contrast, results from Experiment 1 did not fully match results from Experiment 2.1. Like adults in Experiment 2.I, adults tested on Experiment 1 also used acoustics (VOT) as the most impactful source of information in deciding what they heard across both talkers. However, they only integrated contextual information when listening to the non-native talkers. These contrasting results appear to have something to do with the task design, as Experiment 1 only differs from Experiment 2.I in its use of a picture selection task over a match/mismatch task. This suggests other cognitive processes might be relevant in influencing how a listener categorizes phonemes. In fact, the author tested both experiments and felt that it was more challenging to decide whether a picture matched (Experiment 2) than to choose between two pictures (Experiment 1). If the match/mismatch task (Experiment 2) is more cognitively difficult than the picture selection task (Experiment 1), it may cause participants to utilize sentential context more. Additionally, the failure to find an effect of context for the native talkers with the picture selection task (but not with our match/mismatch task) further supports the finding from Experiment 2.I that the effect of context is stronger when listening to non-native talkers, to the extent that the effect for native talkers disappeared.

Our second research question asked if children ages 5-13 years show adult-like weighting of acoustic and contextual cues when listening to native and non-native talkers. Results from Experiment 2.II – a match/mismatch task – suggest that in some ways they do, but in others they do not. Like adults in Experiment 2.I, children weighted acoustics (VOT) most highly, as well as showed a context effect for both native and non-native talkers. On the other hand, they show no difference in their weighting of sentential context for the native vs. the non-native talkers.

These results further our understanding of how children make use of sentential context when listening to non-native talkers. Holt and Bent (2017) found that children ages 5-7 years benefitted from sentential context while repeating native and non-native speech. While performance was overall better for native speech, crucially, the effect of context did not differ between the native and non-native speech. The same use and weighting of sentential context across talkers for children was found in our study. There are now two studies suggesting children may not be impacted by accents in the same way as adults, despite substantial differences in task design (repeating sentences vs. phoneme categorization). Additionally, Holt and Bent (2017) saw an increase of benefit from sentential context for the older children. As our sample size was much smaller, we were unable to look at a comparison across child ages.

Creel, Rojo, and Paullada (2016), on the other hand, did look at the interaction of multiple cues (contextual and visual cues) in preschool-aged children perceiving non-native and native speech via four different experiments. They found that when children were *only* provided sentential context cues or *neither* sentential context *nor* visual cues, they performed better on a task of repeating words when listening to the native vs. non-native speech. In other words, they did not find an effect from sentential context, but did find an effect of talker accent. In contrast, we found a significant effect of sentential context, which did not differ based on whether the talker had a native or non-native accent. This could have to do with the difference in task design, as children in the key experiments in Creel, Rojo, and Paullada (2016) completed a word-repetition task, while children in the current study completed a phoneme categorization task. Another possible reason for the difference in results is the first language of the non-native talkers. Creel, Rojo, and Paullada used Mexican Spanish-accented English. Meanwhile, our study used Mandarin Chinese-accented English. The children Creel and colleagues tested (Californians) may have more exposure to the non-native accent they used than the children we tested (Mississippians) have exposure to the non-native accent we used. Finally, ages of the participants and sample sizes differed between their study and ours.

A. Limitations

There are several limitations to our study. First, phonetic features and phonological patterns vary from language to language and will affect the characteristics of one's non-native English accent. This means the specific native language of the non-native talkers could have influenced how listeners responded, meaning these findings cannot be generalized to all non-native accents of English. Future research could replicate our study with multiple talkers of multiple different accents.

Second, the listener’s experience with the language in question could have influenced how much they relied on sentential context. A few participants had minor exposure to Mandarin Chinese, which potentially could have affected their results. However, we did not believe it was enough exposure to have had a significant impact. Future research could explore a potentially more familiar non-native accent in the United States, such as Spanish-accented English. It is also worth noting that a few of our participants were bilingual from birth (though none spoke Mandarin), and their exposure to and ability to speak multiple languages could also have affected their results. Previous work has shown that even experience with an accent over the course of an experiment leads to less use of contextual information (e.g., Schertz & Hawthorne, 2018). This indicates that both long-term (out of the lab) and short-term (in the lab) experience may be relevant.

Finally, our sample size for children was not very large, and we tested a fairly large age range from 5-13 years. Looking further into a narrower range of ages, or comparatively across a larger number of children covering a range of ages, could allow for more confident claims about the way children perceive non-native speech. Specifically, future research could explore older children's weighting of cues to see when their responses become adult-like, meaning an interaction between talker accent and use of sentential context would be found. Additionally, looking into younger children's weighting of cues would be interesting to see how cue integration differs as age lowers.

B. Conclusion

The current results suggest that, while children and adults both use sentential context when deciding what phoneme they hear in both native and non-native speech, they differ in their weighting of it depending on the talker. That is, adults weight context more highly for non-native talkers, while children weight context similarly across both native and non-native talkers. These finding have implications for the development of phoneme categorization from childhood to adulthood. While deciding what phoneme one hears may seem like a simple automatic task, we see that there are still differences in how children and adults complete this fundamental task. Very few previous studies have investigated the interaction of multiple cues or compared multiple ranges of ages of children listening to non-native speech, so the current results are key in furthering our knowledge of the subject. In addition, results have implications for facilitating better communication between second language learners and native talkers. English language teachers may be able to teach specific strategies to their students to enhance communication, especially for listeners with auditory processing deficits.

LIST OF REFERENCES

Barker, B. A., & Turner, L. M. (2015). Influences of foreign accent on preschoolers' word recognition and story comprehension. *Applied Psycholinguistics, 36*, 1111–1132.

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. C. (2015). lme4: Linear mixed-effects models using Eigen and S4, version 1.1-12, https://cran.r-project.org/web/packages/lme4/index.html (Last viewed March 12, 2018).

Bent, T. (2014). Children's perception of foreign-accented words. *Journal of Child Language, 41*, 1334–1355.

Bent, T., & Atagi, E. (2015). Children's perception of nonnative- accented sentences in noise and quiet. *Journal of the Acoustical Society of America, 138*, 3985–3893.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America, 103*(5), 2670–2676.

Bradlow, A., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729.

Clarkson, R. L., Eimas, P. D., & Marean, G. C. (1989). Speech perception in children with histories of recurrent otitis media. *Journal of the Acoustical Society of America, 85*, 926-33.

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 291-99.

Creel, S. C., Rojo, D. P., & Paullada, A. N. (2016). Effects of contextual support on preschoolers' accented speech comprehension. *Journal of Experimental Child Psychology, 146*, 156-180.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology, 3*(479), 1-15.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing Research, 229*, 132-147.

De Rosario-Martinez, H. (2015). Package 'phia,' version 0.2-1, https://github.com/heliosdrm/phia (Last viewed March 12, 2018).

Deterding, D., & Nolan, F. (2007). "Aspiration and voicing of Chinese and English plosives." *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 385–388.

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

Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *Journal of the Acoustical Society of America, 108*, 3023–3029.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *Journal of the Acoustical Society of America, 111*, 2242–2249.

Flege, J. E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics, 15*, 67–83.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 462*, 502-504.

Haggard, M. P., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America, 47*(2B), 613-617.

Holt, R. F., & Bent, T. (2017). Children's use of semantic context in perception of foreign-accented speech. *Journal of Speech, Language, and Hearing Research, 60*, 223–230.

Holt, L., & Lotto, A. (2010). Speech perception as categorization. *Attention, Perception & Psychophysics, 72*, 1218–1227.

Holtby, A. K. (2010). *The Effect of Age and Exposure in the Development of L2 Accent Perception (PMC Working Paper series, 10-01)*. Edmonton, CA: Prairie Metropolis Centre.

McClelland, J., Mirman, D., & Holt, L. (2006) Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*, 363–369.

McDonald, M., Gross, M., Buac, M., Batko, M., & Kaushanskaya, M. (2017). Processing and comprehension of accented speech by monolingual and bilingual children. *Language Learning and Development, 14*(2), 113-129.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97.

R Core Team (2018). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/ (Last viewed March 12, 2018).

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America, 51*, 1296-1303.

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

Ryan, C. (2013). Language use in the United States: 2011 (Report No. ACS-22). *U.S. Census Bureau*, 1-16.

Schertz, J., & Hawthorne, K. (2018). The effect of sentential context on phonetic categorization is modulated by talker accent and exposure. *Journal of the Acoustical Society of America, 143*(3), EL231-EL236.

Strand, E. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*, 86-99.

Whalen, D. H., Abramson, A. S., Lisker, L., & Maria, M. (1993). *F*0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America, 93*(4), 2152-2159.

Yost, W. A. (1992). Auditory image perception and analysis. *Hearing Research, 56*, 8–19.

APPENDIX A: *Adult Language Background Questionnaire*

The purpose of this questionnaire is to learn something about your language history: which languages you know and which languages you hear on a regular basis.

1. Age:
2. Gender: M / F
3. As far as you know, is your hearing normal? (If no, please explain)
4. As far as you know, is your vision normal or corrected-to-normal? (If no, please explain.)
5. Have you ever been diagnosed with a speech, language, or reading problem? (If yes, please explain.)
6. Where were you born? City: State/Province: Country:
7. Please list any places **outside of North America** that you have lived for over three months, and indicate how long and at what age you lived there:

Location (city and country)	Age	Length of time
e.g. Melbourne, Australia	*e.g. 6-8 years old*	*e.g. 2 years*

8. Please answer the following questions about the language(s) your parents speak (or other adults in your household while you were growing up).

	Father	Mother
First language(s)		
Language(s) they speak to you (give percentages if >1)		

9. For each language listed above, indicate the average percentage of time you use each language currently (these percentages should add up to a total of 100%).

Language:					
Percentage:					

10. For each language that you speak or have studied, please answer the following questions:

Language:	**1. English**	**2.**	**3.**	**4.**
Did you learn this language at home or at school?				
At what age were you first exposed to this language?				
At what age did you feel comfortable speaking this language? (write "not yet" if it is not the case)				
Do you think you have a non-native accent in this language?				
Who do you most often use this language with? (e.g. family, friends, professors)				
On a scale from 1 (beginning learner) to 10 (completely native), how do you rate your ability to **speak** this language?				
On a scale from 1 (beginning learner) to 10 (completely native), how do your rate your ability to **understand** this language?				

11. For some studies we are particularly interested in whether you have experience with people who have particular types of foreign/regional accents or different kinds of language disorders.

    Do you ever ever interact with someone whose native language is Mandarin Chinese?
    ___ Yes ___ No

    If yes, how often and in what context? (e.g., professors in the classroom; Grandparent on the telephone 1 hour per week)

APPENDIX B: *Parent Language Background Questionnaire*

Instructions: The primary caregiver should fill out the following survey. If a question does not apply please write "N/A" for not applicable. If you are not sure, please write "not sure."

Your relation to the child: __________________________ Today's date: ____________

Child's birthdate: _________________________________ Child's age: ____________

Child's place of birth (City, State, Country): ____________________________________

Child's gender (please check one): ___ Male ___ Female

Child's race (please check one):

___ American Indian/Alaska Native ___ Black or African American

___ Asian ___ White

___ Native Hawaiian/Other Pacific Islander ___ More than one race (list below)

___ Other: ___________________________ _________________________

Your race (if different from your child's): ______________________________________

Child's ethnicity (please check one): ___ Hispanic/ Latino ___ Not Hispanic/ Latino

Your ethnicity (if different from child's): ___ Hispanic/ Latino ___ Not Hispanic/ Latino

What is your highest level of education?

___ some high school ___ college degree

___ high school diploma ___ graduate degree

___ some college or vocational training

Hearing & Vision

1. Does your child have any vision impairments (with or without glasses)? __ Yes __ No
If yes, how old was your child when you first realized he/she had vision impairment? __

2. Does your child wear corrective lenses or glasses? __ Yes __ No
If yes, how old was your child when he/she received corrective lenses or glasses? __

3. Does your child have any hearing loss? __ Yes __ No
If yes, how old was your child when you first realized he/she had hearing loss? __

4. Does your child wear a hearing aid? __ Yes __ No
If yes, how old was your child when he/she received the hearing aid? __

5. Has your child ever been diagnosed with a speech or language problem? __ Yes __ No
If yes, how old was your child when he/she received the diagnosis? __

6. Has your child ever been treated for a speech or language problem? __ Yes __ No
If yes, how old was your child when he/she received the hearing aid[1]? __

Language

1. Does your child speak English as his/her first language? __ Yes __ No

2. Does your child regularly hear any other languages? __ Yes __ No

    If yes, please fill out the following:

    Language: *English*
    How old was your child when they were exposed to English? ______
    How many hours per week is your child exposed to English? ______
    Who speaks English to your child (mother, father, grandparent, teacher, etc.)?
    ____

    Language: __________________
    How old was your child when they were exposed to this language? ______
    How many hours per week is your child exposed to this language? ______
    Who speaks English to your child (mother, father, grandparent, teacher, etc.)?
    ____

3. For each language listed above, indicate the average percentage of time your child *uses* the language. (These percentages should add up to a total of 100%.)

Language: English	Language:	Language:
Percentage:	Percentage:	Percentage:

4. For some studies we are particularly interested in whether your child has experience with people who have different types of foreign/regional accents or different kinds of language disorders.

    Does your child ever interact with someone whose native language is Mandarin?
    ___ Yes ___ No

    If yes, please describe.

[1] *We later realized the subquestion of Question 6 under the "Hearing & Vision" section is incorrect. It should ask, "If yes, how old was your child when he/she was treated for the speech or language problem?" This question was only relevant for two parents, whom we clarified the information for as necessary.*