

Published in final edited form as:

J Exp Child Psychol. 2013 November ; 116(3): 728–737. doi:10.1016/j.jecp.2013.05.008.

Spectral information in nonspeech contexts influences children's categorization of ambiguous speech sounds

Daniel G. Hufnagle, Lori L. Holt, and Erik D. Thiessen*

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

For both adults and children, acoustic context plays an important role in speech perception. For adults, both speech and nonspeech acoustic contexts influence perception of subsequent speech items, consistent with the argument that effects of context are due to domain-general auditory processes. However, prior research examining the effects of context on children's speech perception have focused on speech contexts; nonspeech contexts have not been explored previously. To better understand the developmental progression of children's use of contexts in speech perception and the mechanisms underlying that development, we created a novel experimental paradigm testing 5-year-old children's speech perception in several acoustic contexts. The results demonstrated that nonspeech context influences children's speech perception, consistent with claims that context effects arise from general auditory system properties rather than speech-specific mechanisms. This supports theoretical accounts of language development suggesting that domain-general processes play a role across the lifespan.

Keywords

Speech perception; Categorical perception; Language development; Phonemes; Acoustics; Phonetic categories

Introduction

Acquiring native language phonetic categories presents a complex learning problem because mappings from acoustics to phonetic categories are not straightforward. Young learners must discover acoustic dimensions that are relevant to distinguishing native phonetic categories and learn how variability along these dimensions relates to categories. Learners accomplish this despite much acoustic variability; acoustic dimensions signaling phonetic categories vary radically across talkers (Lieberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Nygaard, Sommers, & Pisoni, 1994; Strand & Johnson, 1996) and within talkers (Hillenbrand, Getty, Clark, & Wheeler, 1995; Lieberman, 1957; Lisker & Abramson, 1967). Adult listeners rely on context to deal with acoustic variability (see Repp, 1982, for a review). The mechanisms underlying the influence of context on adult speech categorization have been well investigated (e.g., Holt & Lotto, 2002; Lotto & Holt, 2006). Context is likely to be important for language learners as well; for example, infants and young children are sensitive to the lexical context in which phonemes occur (e.g., Swingley, 2009; Thiessen & Yee, 2010). To better understand the developmental progression of children's use of contexts in speech perception and the mechanisms underlying that development, we created a novel experimental paradigm testing 5-year-old children's speech perception in several

acoustic contexts. This paradigm allows us to test the theoretical claim that context effects in speech perception arise from general auditory processing rather than speech-specific mechanisms.

Context can be powerful in resolving perceptual ambiguity. For example, when perceptually ambiguous sounds from a /d/-to-/t/ series are followed by *-ash*, making a series that varies from *dash* to *tash*, people are more likely to categorize tokens as /d/ (consistent with the real word *dash*), whereas the same tokens are more often categorized as /t/ in *-ask* contexts that vary perceptually from *dask* to the real word *task* (Ganong, 1980). Visual context also disambiguates acoustics; watching faces articulate /aba/ or /ada/ shifts categorization of perceptually ambiguous acoustic tokens to be consistent with visual faces (Bertelson, Vroomen, & de Gelder, 2003). Moreover, phonetic context also influences speech categorization; when sounds from a /ga/-to-/da/ series are preceded by /al/, they are more often perceived as /ga/, whereas sounds that are preceded by /ar/ are more often perceived as /da/ (Mann, 1980).

Context sounds need not be speech in order to influence categorization of speech targets. Nonlinguistic tones that mimic critical spectral characteristics of /al/ and /ar/ shift perception of a /ga/-to-/da/ continuum similarly (Fowler, Brown, & Mann, 2000; Holt & Lotto, 2002; Lotto & Holt, 2006; Lotto & Kluender, 1998; Lotto, Sullivan, & Holt, 2003). Thus, it seems that general auditory characteristics of context, not specific to the speech signal, can influence speech categorization. In the example above, the /al/ and nonspeech tones that model it possess relatively higher acoustic frequencies and shift perception of the speech targets toward /ga/, the response alternative possessing lower frequency information. Conversely, speech (/ar/) or tone contexts with relatively lower frequencies shift perception toward /da/, the higher frequency response alternative. These spectrally contrastive effects have been replicated across adult categorization of vowels (Holt, Lotto, & Kluender, 2000), lexical tones (Huang & Holt, 2009, 2011), and other consonants (Holt, 1999; see Lotto & Holt, 2006, for a review). In each case, speech categorization is robustly influenced contrastively by the frequencies of sounds adjacent to the targets, and this influence appears to originate from very general properties of auditory processing not specific to speech (Lotto, Kluender, & Holt, 1997).

The claim that context effects arise from domain-general auditory processes is consistent with theoretical accounts that rely on domain-general mechanisms to explain language development (e.g., Saffran & Thiessen, 2007). However, it is as yet unclear whether these domain-general mechanisms play a role in context effects in speech perception for infants and children. Although context effects on speech perception have been demonstrated during infancy (Fowler, Best, & McRoberts, 1990), such demonstrations have focused on speech contexts. The impact of nonspeech contexts on speech categorization has been tested only in adults (e.g., Holt, 2005, 2006a; Lotto & Kluender, 1998). Discovering whether children are similarly influenced by nonspeech contexts is important for understanding the contribution of these domain-general processes to language development.

These processes are a plausible contributor to language development only if they are available to young language learners, but children might not be able to use nonlinguistic context like adults. One possibility is that children lack the cognitive capacity necessary to use long-term acoustic contexts such as those that evolved over several seconds of acoustic input in the Holt (2005, 2006a) studies. Children may lack memory resources necessary to store spectral information gained from context in order to use it by the time the to-be-interpreted signal arrives. Alternatively, the effect of nonlinguistic contexts on speech categorization may depend on strategies developed as a consequence of expertise in speech perception rather than as a consequence of domain-general perceptual processing. That is, if

listeners become better at detecting general spectral patterns as they gain experience in exploiting speech contexts to aid in speech categorization, this speech-centric ability may generalize to other acoustic signals. Under this scenario, well-developed speech categories would be a prerequisite for observing nonlinguistic acoustic context effects because these effects depend on good identification of phonetic categories.

Conversely, the fact that children do not yet possess adult-like phonetic categories may actually amplify the effects of context. Because young language learners are still discovering their language's phonetic categories, they sometimes respond to sounds that are phonemically contrastive as though they are ambiguous (e.g., Stager & Werker, 1997). The effects of context are exaggerated for perceptually ambiguous speech tokens, so context may play an especially large role in the speech perception of young language learners. The nature of context-dependent speech perception is such that perceptually ambiguous speech sounds tend to be shifted toward one category alternative or another, thereby resolving some ambiguity. Domain-general mechanisms that sharpen perception of ambiguous speech in adults may also shape language development. During acquisition, a process that uses contextual information available in the ongoing acoustic stream to guide interpretation speech sounds may alleviate some effects of variability in the acoustic speech signal. For these domain-general mechanisms to inform language acquisition, however, children must be able to take advantage of them while the process of language acquisition is ongoing.

The domain generality of the processes involved in speech perception and language acquisition remains the focus of a wide variety of research (e.g., Holt & Lotto, 2010; Lidz, Gleitman, & Gleitman, 2003; Thiessen, 2011). Assessing whether general perceptual processes that use spectral contrast support speech categorization in children, as they do in adult listeners, is a novel question that gets to the heart of whether acquiring speech categories relies on general perceptual systems or a system dedicated to speech. This experiment modifies traditional psychoacoustic methods, reducing the cognitive load and tedium of traditional methods. The experiment involved two sessions. The first session established, for each child, which sounds in a /ga/-/da/ series were most ambiguous without contexts, and the second one presented those ambiguous sounds with tone contexts. Evidence of a context effect caused by nonlinguistic tonal contexts would suggest a domain-general perceptual mechanism that may help to explain how children appear to perceive speech better in real-world communication than they do in standard laboratory measures of phonetic categorization.

Method

Participants

The participants were 17 5-year-old children (average age = 5.5 years, range = 5.2–6.1, 7 girls and 10 boys) from the Children's School at Carnegie Mellon University. Children were tested individually in private rooms at school. All children spoke English fluently as their dominant language and clearly articulated /da/ and /ga/ outside of the context of the experimental task (as judged by the first author, who taught English as a second language for many years). No child had apparent hearing problems or language disorders. In addition, 11 adult native English speakers with no reported hearing or language problems from the Carnegie Mellon University community were also tested for comparison on the task used in the second session. Adults were tested individually in sound-attenuated booths. (Note that groups of younger children—3 and 4 years of age—were also tested, but interpreting whether their data indicated difficulty with the task or difficulty in perception was hard. Therefore, their data are not discussed here.)

Procedure

Session 1 took 12 to 15 min. Children heard a story, answered comprehension questions, and performed an identification task in that order. All children answered story-related comprehension questions well (indicating that they heard the difference between /da/ and /ga/ in the story), so these questions are not discussed further. In Session 2, which took approximately 10 min, children listened to a story and performed an identification task. Children wore headphones during both identification tasks, where stimuli were presented diotically at comfortable listening levels via custom software. Adults performed only the first identification task, entering responses themselves.

The current method is child-friendly, incorporating traditional two-alternative forced-choice identification tasks into a game, reducing demands on children, and allowing them to complete the tasks easily. Each session presented /da/ and /ga/ as different characters or objects in a story in which children heard the experimenter say /da/ and /ga/ in meaningful situations. In Session 1, the experimenter repeated each sound nine times in the story and asked six questions that required children to differentiate the sounds successfully. In Session 2, /da/ and /ga/ were given different referents in an entirely different story.

Stimuli

Speech targets were derived from natural /ga/ and /da/ recordings (20-kHz sampling rate, 16-bit resolution) of a male native English speaker and have been used in previous research (Holt, 2005, 2006a, 2006b; Laing, Liu, Lotto, & Holt, 2012; Wade & Holt, 2005). One /ga/ and one /da/ token that were nearly identical in spectral and temporal properties, except for onset frequencies of the second (F2) and third (F3) formants, served as endpoints in constructing the nine-step continuum. F2 and F3 are the primary acoustic differences between /ga/ and /da/ (Lotto & Kluender, 1998; Mann, 1980). Linear predictive coding (LPC) analysis on endpoint tokens determined a filter series that spanned endpoints (Analysis-Synthesis Laboratory, Kay Elemetrics). F2 and F3 onset frequencies varied approximately linearly between /ga/ and /da/ endpoints in resulting filters, which were excited by the LPC residual of the original /ga/ production to create an acoustic series spanning natural endpoints in approximately equal steps, with intermediate, perceptually ambiguous tokens that sounded very natural. Targets matched the /da/ endpoint in root mean square (RMS) energy. Targets were 589 ms long. Experiments involving children are strictly time limited in the testing environment that we used, so we were limited to using only seven of the nine targets in order to fall within the time limits of the experiment. Adults' category boundaries generally occur between Tokens 5 and 8 (see Holt, 2005, 2006a, 2006b; Wade & Holt, 2005), so Tokens 2 and 4 were removed in order to encourage response patterns in which both response choices are represented more or less equally (see Fig. 1). Endpoint /ga/ and /da/ stimuli are natural utterances spoken by a real talker.

Identification task

In both sessions, children were told that they would hear /da/ or /ga/ through headphones; the experimenter was unable to hear the sound presented over the headphones. Children said the sounds to the experimenter, who entered responses. Stimuli were not repeated; if children were unsure or if noise interfered, trials were not counted (15 of 1586 trials [$>1\%$] were not counted). If children responded inaudibly, the experimenter asked for clarification. Children whose responses did not vary over approximately 10 trials were reminded to use both choices. In the first session, the first 10 trials were practice trials for /ga/ and /da/ endpoint targets in order to test identification accuracy of unambiguous endpoint stimuli without potential interference from ambiguous mid-series stimuli. Familiarizing listeners with unambiguous stimuli is common in speech categorization research (e.g., Mann, 1980). Trials 1 to 4 were presented in a predetermined order, giving equal presentation of endpoints

and allowing corrective feedback. An additional 70 trials were ordered randomly without feedback and immediately followed unambiguous trials, consisting of 10 repetitions each of /ga/ and /da/ endpoints and stimuli 3, 5, 6, 7, and 8 from the continuum. Children were not told that sounds would be ambiguous.

Trials in the second session consisted of a nonspeech context—a series of sine wave tones designed to create peaks of spectral energy at specific frequencies in the long-term average spectrum of contexts—and a speech target. Context affects identification of perceptually ambiguous targets more than unambiguous targets (Holt, 2005, 2006a), so the most ambiguous pair of stimuli for each child was carefully identified from Session 1 response data. Children's ambiguous pairs were determined by finding targets for which identification judgments reliably shifted from more than 50% /ga/ to more than 50% /da/ and consisted of adjacent targets on the continuum. For children selected for Session 2, when judgments reached the 50% threshold, they remained at or below 50% through the rest of the continuum. Target pairs included all possible non-endpoint stimulus pairs (3 + 5, 5 + 6, 6 + 7, and 7 + 8). Of the 17 children who participated in the first session, 13 met the criteria to participate in the second session.

Nonspeech contexts consisted of 22 sine wave tones presented across 2.2 s (see Fig. 2). The first 21 tones in low-frequency contexts varied between 1500 and 2100 Hz in 100-ms steps (7 tones presented three times each, $M = 1800$ Hz), creating a spectral distribution that is expected, from adult studies (e.g., Holt, 2005), to shift categorization toward /da/, with high-frequency contexts varying between 2500 and 3100 Hz ($M = 2800$ Hz), creating a distribution predicted to shift categorization toward /ga/. Middle-frequency contexts sampled frequencies between 2000 and 2600 Hz ($M = 2300$ Hz); these were not expected to affect categorization and were used for endpoint targets only. Tones were matched in RMS energy to speech targets. Mean frequencies and other general context parameters followed those used by Holt (2005), with differences limited to frequency range for contexts (600 Hz here vs. 1000 Hz for Holt, 2005) and the step size that was used (100-Hz steps here vs. 50-Hz steps for Holt, 2005). The final tone was always 2300 Hz, which does not shift adults' identification responses relative to responses to speech targets presented without contexts (Holt, 2005); this ensured that any observed shift in children's identification required influence from some average of spectral characteristics of contexts rather than the nearest tone alone, thereby implicating the use of higher level processes rather than earlier processes of the peripheral auditory system. Conditions were defined in terms of the spectral distribution sampled by tones, with 10 random orderings of high- and low-tone contexts and 6 random orderings of medium-tone contexts being presented to listeners.

Session 2 consisted of 52 trials: 6 medium-context, nonbiasing /ga/ endpoint trials, 6 medium-context /da/ endpoint trials, 20 low-context, /da/-biasing ambiguous target trials (10 for each target from the child's most ambiguous pair), and 20 high-context, /ga/-biasing ambiguous target trials. Unambiguous trials ensured that children reliably identified at least some trials because other targets were chosen to be perceptually ambiguous. An initial 4 trials presented two repetitions of unambiguous /da/ and /ga/ targets in a predetermined order to allow for feedback.

Results and discussion

Session 1

Responses to perceptually unambiguous /ga/ and /da/ endpoints were 82% and 89% correct, respectively (see Fig. 3). As expected, children's responses were not as categorical as adults' responses, and children's identification curves did not resemble the shape of categorical perception curves (see Harnad, 1987). Identifying category boundaries often involves fitting

idealized curves to response data in order to model inflection points of response functions (e.g., Finney, 1947). However, such analyses consistently produced errors for children's response data, failing to find category boundaries. To quantify linearity of children's responses, we compared slopes of the straight lines between each response curve's endpoints with the slope of its intermediate points (Stimuli 5, 6, and 7; see Fig. 3). Slopes should be different in nonlinear curves, with steeper slopes at intermediate stimuli than the shallower slope between the endpoints. For children, slopes were not different, $t(16) = 0.52, p = .61$. Conversely, adults' slopes differed significantly, $t(10) = 2.54, p = .029, \eta^2 = .39$. That is, children yielded response

Session 2

Of the 17 children, 13 met criteria for determining their most ambiguous target pair and participated in Session 2. Bearing out our predictions, children categorized identical ambiguous targets differently depending on contexts, with responses shifting 37% in the predicted directions (see Fig. 4). Children more often indicated *ga* in response to the ambiguous speech targets when tone contexts had high mean frequencies ($M = 80\%$) than when those same speech targets had contexts with low mean frequencies ($M = 43\%$), $t(12) = 6.57, p < .001, \eta^2 = .80$ (all t tests were Bonferroni corrected). Critically, children's categorization of ambiguous speech targets was influenced robustly by the distribution of acoustic frequencies in contexts. Like adults, directionality of influence was spectrally contrastive; contexts sampling higher frequencies increased *ga* responses, the target with lower frequency spectral energy, whereas tones sampling lower frequencies increased *da* responses, which has a higher frequency spectrum. These findings support our hypothesis that children can use spectral information present in nonspeech contexts to inform categorization.

A repeated measures analysis of variance (ANOVA) with four levels of sound type (medium-context /*ga*/, high-context ambiguous, low-context ambiguous, and medium-context /*da*/) with percentage *ga* responses as the dependent variable revealed a main effect of sound type, $F(3, 36) = 90.7, p < .001, \eta^2 = .61$ (see Fig. 4). Children's categorization of perceptually ambiguous stimuli preceded by high tones was statistically indistinguishable from categorization of /*ga*/ endpoint stimuli in neutral context, $t(12) = 2.33, p = .028$ (Bonferroni corrected). This effect was not symmetrical with the /*da*/ endpoint, $t(12) = 8.115, p < .001, \eta^2 = .84$ (Bonferroni corrected); that is, children categorized the endpoint /*da*/ in neutral tone context more often as *da* than the ambiguous tokens in the context of lower frequency preceding tones. Nonetheless, the net effect was to sharpen children's categorization judgments of ambiguous tokens. Perceptually ambiguous sounds were categorized differently in high context than targets categorized in isolation, $t(12) = 5.80, p < .001, \eta^2 = .70$ (Bonferroni corrected), but responses to ambiguous targets preceded by low contexts and no contexts did not differ significantly, although they shifted nominally in the predicted direction, $t(12) = 1.40, p = .20$. Note that high contexts produce larger effects than low contexts in adult listeners as well (see Holt, 2006a). Overall, 12 of 13 children shifted their responses in the predicted direction; the remaining child's categorization was unchanged between Sessions 1 and 2. Paired t tests revealed no significant difference in endpoint target categorization in isolation versus neutral tone contexts, $t(12) = 0.05, p = .96$.

Developing phonetic categories are refined well into middle childhood (e.g., Elliott, Hammer, & Scholl, 1990; Hazan & Barrett, 2000; Nittrouer, 2002; Nittrouer & Miller, 1997; Ohde & Haley, 1997; Simon & Fourcin, 1978; see Zevin, 2012, for a review). The fact that children's categories are developing may explain why children show difficulty in tasks that require attention to slight phonemic differences (e.g., Stager & Werker, 1997; Swingley, 2007; Thiessen, 2011). Our results address a pragmatic question: How can children produce and comprehend speech in natural linguistic environments when phonetic categories are less

than fully developed? We suggest that children take advantage of acoustic context to resolve perceptual ambiguities. Here, children's speech identification responses were influenced strongly by distributionally defined spectral information present in a sequence of nonspeech tones.

These results provide insight into mechanisms responsible for listeners' sensitivity to contextual information. Long-term spectral contrast effects in children demonstrate that children have memory processes necessary to use spectral information to inform perception. This information unfolds over a time frame of several seconds. Whether children respond to all or part of the context is a topic to be explored further. The mechanism emphasizes the contrast between the spectral information in the context and the target to sharpen categorization of perceptually ambiguous speech. These findings also support theoretical claims that general auditory mechanisms play a role in the development of speech perception. Children are sensitive to nonlinguistic context, in addition to linguistic context (Fowler et al., 1990), when resolving ambiguous phonetic exemplars. The fact that nonspeech contexts affect children's speech categorization of tokens not yet perceived very categorically supports our claim that long-term spectral contrast effects are due to general auditory mechanisms that are an integral part of the auditory system and do not require adult-like phonetic categories. That is, the nonspeech context effect cannot result from a speech-specific mechanism requiring great expertise with speech categories. Such a mechanism would require enough expertise to allow the mechanism to generalize beyond speech to include nonspeech contexts, and these children are demonstrably not experts at identifying syllables from these categories. The presence of these effects in young listeners whose categories are still developing allows for the possibility that auditory mechanisms use spectral context information to aid development of phonetic categories. Combined with other contributions to phonetic categorization, this mechanism could help learners to discover the detailed structure of phonetic categories, thereby allowing increasingly categorical responses to context-free ambiguous speech sounds as development progresses.

Our work touches on an important developmental question: How do children's categories reach an adult-like state? We believe that sensitivity to context plays an important role. The current results demonstrate that *acoustic* context shapes children's phonetic perception. Among adult listeners, nonspeech tones that model the long-term average spectrum of a particular voice shift speech perception in a manner like the voice they model, suggesting that general auditory mechanisms may play a role in talker normalization (Holt, 2005; Huang & Holt, 2012; Laing et al., 2012). From the current results, we can conclude that such mechanisms are available to support phonetic development and speculate that they may help children to resolve acoustic variability arising from different talkers.

Likewise, it may be the case that experience with phonetic categories in the context of lexical forms helps to refine children's phonetic representations. For example, knowing more words can lead to better discrimination between similar word forms (e.g., *bih* and *dih*) in word-object association tasks (e.g., Werker, Fennell, Corcoran, & Stager, 2002). The growth of the lexicon is likely to contribute to emerging phonetic categories (Swingley, 2009). Exposure to sounds in distinct contexts facilitates use of the phonemic distinctions. Infants who cannot detect a switch from *daw* to *taw* in a word-object association task can do so when they are preexposed to the sounds in differentiating contexts *dawbow* and *tawgoo*, but not when they are preexposed to the sounds in non-differentiating contexts *dawbow* and *tawbow* (Thiessen, 2007, 2011; Thiessen & Yee, 2010). Learning words may provide important contextual support that guides children toward more adult-like phonetic categories.

If context informs children's perception, there is a ready explanation for why children inconsistently categorize sounds in standard speech identification tasks: Sounds are presented in isolation. Without context, which is ubiquitous outside the laboratory, young listeners may lack cues vital to identifying speech consistently. This may help to account for the apparent inconsistency between infants' rapid speech category development and the relatively protracted continuing development that lingers through childhood (see Zevin, 2012). The methodological differences in determining speech perception abilities that are required for work with a developing population may explain the apparent differences between infants and older children in two ways. First, infant methods commonly use tasks that are easier than the standard adult psychophysical tasks, relying on prototypical unambiguous targets that are presented in infant-directed speech, which can exaggerate important cues to speech categories (Fernald, 1984; Snow, 1977; Werker & McLeod, 1989). Many tasks commonly used with very young listeners measure infants' ability to detect change rather than the overt categorization that is required of children who are asked to label speech sounds, leaving open the possibility that infants are detecting some change that is not, for them, a change in category. Second, standard adult methods used for older children may be more difficult for children because of their lack of context.

The fact that children are sensitive to the distribution of a sequence of tones, and that this distributional information influences their performance in a speech perception task, is consistent with infants' sensitivity to statistical information originating from other sources (e.g., Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996). This domain-general sensitivity to statistical information allows infants and children to discover structure in language and detect which features of input are relevant (e.g., Thiessen & Saffran, 2007). Identifying the kinds of statistical information that influence children can inform theories about the processes and representations that are available to learners. Here we have demonstrated that children can, at a minimum, determine some distributionally defined spectra of a series of tones that builds up over 2.3 s of context and can use that context to categorize speech sounds less ambiguously.

References

- Bertelson P, Vroomen J, de Gelder B. Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*. 2003; 14:592–597. [PubMed: 14629691]
- Elliott LL, Hammer MA, Scholl ME. Fine-grained auditory discrimination and performance on tests of receptive vocabulary and receptive language. *Annals of Dyslexia*. 1990; 40:170–179.
- Fernald, A. The perceptual and affective salience of mothers' speech to infants. In: Feagans, L.; Garvey, C.; Golinkoff, R., editors. *The Origins and Growth of Communication*. Norwood, NJ: Ablex; 1984.
- Finney, DJ. *Probit analysis*. Cambridge, UK: Cambridge University Press; 1947.
- Fowler CA, Best CT, McRoberts GW. Young infants' perception of liquid co-articulatory influences on following stop consonants. *Perception and Psychophysics*. 1990; 48:559–570. [PubMed: 2270188]
- Fowler CA, Brown JM, Mann VA. Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26:877–888. [PubMed: 10883999]
- Ganong WF. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*. 1980; 6:110–125. [PubMed: 6444985]
- Harnad, S. *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press; 1987.
- Hazan V, Barrett S. The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*. 2000; 28:296–377.

- Hillenbrand JM, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*. 1995; 97:3099–3111. [PubMed: 7759650]
- Holt LL. unpublished doctoral dissertation. University of Wisconsin–Madison; 1999. Auditory constraints on speech perception: An examination of spectral contrast.
- Holt LL. Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychological Science*. 2005; 16:305–312. [PubMed: 15828978]
- Holt LL. The mean matters: Effects of statistically-defined non-speech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*. 2006a; 120:2801–2817. [PubMed: 17091133]
- Holt LL. Speech categorization in context: Joint effects of non-speech and speech precursors. *Journal of the Acoustical Society of America*. 2006b; 119:4016–4026. [PubMed: 16838544]
- Holt LL, Lotto AJ. Behavioral examinations of the neural mechanisms of speech context effects. *Hearing Research*. 2002; 167:156–169. [PubMed: 12117538]
- Holt LL, Lotto AJ. Speech perception as categorization. *Attention, Perception, & Psychophysics*. 2010; 72:1218–1227.
- Holt LL, Lotto AJ, Kluender KR. Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*. 2000; 108:710–722. [PubMed: 10955638]
- Huang J, Holt LL. General perceptual contributions to lexical tone normalization. *Journal of the Acoustical Society of America*. 2009; 125:3983–3994. [PubMed: 19507980]
- Huang J, Holt LL. Evidence for the central origin of lexical tone normalization. *Journal of the Acoustical Society of America*. 2011; 129:1145–1148. [PubMed: 21428475]
- Huang J, Holt LL. Listening for the norm: Adaptive coding in speech categorization. *Frontiers in Perception Science*. 2012; 3:10.
- Laing EJ, Liu R, Lotto AJ, Holt LL. Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology*. 2012; 3:203. [PubMed: 22737140]
- Lieberman AM. Some results of research on speech perception. *Journal of the Acoustical Society of America*. 1957; 29:117–123.
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- Lidz J, Gleitman H, Gleitman L. Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*. 2003; 87:151–178. [PubMed: 12684198]
- Lisker L, Abramson AS. Some effects of context on voice onset time in English stops. *Language and Speech*. 1967; 10:1–28. [PubMed: 6044530]
- Lotto AJ, Holt LL. Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*. 2006; 68:178–183. [PubMed: 16773891]
- Lotto AJ, Kluender KR. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*. 1998; 60:602–619. [PubMed: 9628993]
- Lotto AJ, Kluender KR, Holt LL. Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*. 1997; 102:1134–1140. [PubMed: 9265760]
- Lotto AJ, Sullivan SC, Holt LL. Central locus for non-speech effects on phonetic identification. *Journal of the Acoustical Society of America*. 2003; 113:53–56. [PubMed: 12558245]
- Mann VA. Influence of preceding liquid on stop consonant perception. *Perception and Psychophysics*. 1980; 28:407–412. [PubMed: 7208250]
- Maye J, Werker JF, Gerken LA. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*. 2002; 82:B101–B111. [PubMed: 11747867]
- Nittrouer S. Specifying voicing differences in children's productions of syllable-final stops: Knowledge vs. skill. *Journal of the Acoustical Society of America*. 2002; 111:S2482. [abstract].
- Nittrouer S, Miller ME. Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*. 1997; 101:2253–2266. [PubMed: 9104027]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychological Science*. 1994; 5:42–46. [PubMed: 21526138]

- Ohde RN, Haley KL. Stop consonant and vowel perception in 3- and 4-year-old children. *Journal of the Acoustical Society of America*. 1997; 102:3711–3722. [PubMed: 9407663]
- Repp BH. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*. 1982; 92:81–110. [PubMed: 7134330]
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–1928. [PubMed: 8943209]
- Saffran, JR.; Thiessen, ED. Domain-general learning capacities. In: Hoff, E.; Schatz, M., editors. *Blackwell handbook of language development*. Oxford, UK: Blackwell; 2007. p. 68-86.
- Simon C, Fourcin A. Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*. 1978; 63:925–935.
- Snow CE. The development of conversation between mothers and babies. *Journal of Child Language*. 1977; 4:1–22.
- Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*. 1997; 388:381–382. [PubMed: 9237755]
- Strand, EA.; Johnson, K. Gradient and visual speaker normalization in the perception of fricatives. In: Gibbon, D., editor. *Natural language processing and speech technology: Results of the 3rd KONVENS Conference, Bielefeld, October 1996*. Berlin: Mouton; 1996. p. 14-26.
- Swingle D. Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*. 2007; 43:454–464. [PubMed: 17352552]
- Swingle D. Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*. 2009; 364:3617–3622.
- Thiessen ED. The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*. 2007; 56:16–34.
- Thiessen ED. When variability matters more than meaning: The effect of lexical forms on use of phonemic contrasts. *Developmental Psychology*. 2011; 47:1448–1458. [PubMed: 21744949]
- Thiessen ED, Saffran JR. Learning to learn: Acquisition of stress-based strategies for word segmentation. *Language Learning and Development*. 2007; 3:75–102.
- Thiessen ED, Yee MN. Dogs, bogs, labs, and lads: What phonemic generalizations indicate about the nature of children's early word-form representations. *Child Development*. 2010; 81:1287–1303. [PubMed: 20636696]
- Wade T, Holt LL. Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of the Acoustical Society of America*. 2005; 118:2618–2633. [PubMed: 16266182]
- Werker JF, Fennell CT, Corcoran KM, Stager CL. Infants' ability to learn phonetically similar words: Effects of age and vocabulary. *Infancy*. 2002; 3:1–30.
- Werker JF, McLeod PJ. Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*. 1989; 43:230–246. [PubMed: 2486497]
- Zevin JD. A sensitive period for shibboleths: The long tail and changing goals of speech perception over the course of development. *Developmental Psychobiology*. 2012; 54:632–642. [PubMed: 22714710]

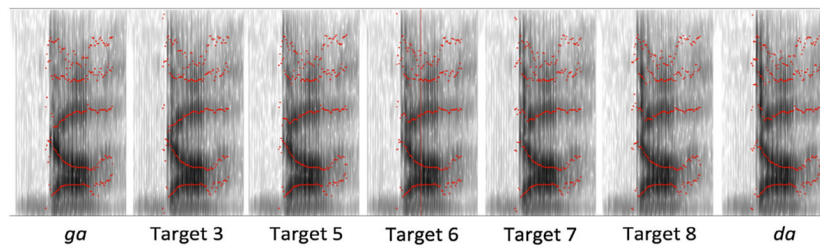


Fig. 1. Spectrograms of speech targets plotted on time by frequency axes. The lines highlight the formant frequencies of the speech tokens.

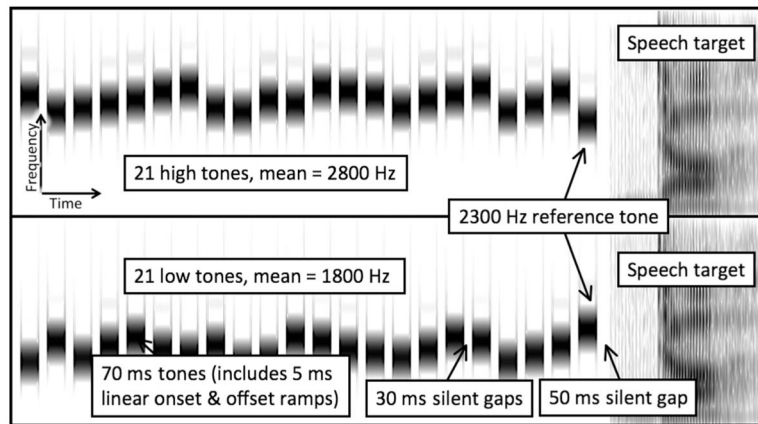


Fig. 2. Representative stimuli. The top panel shows a trial with high-tone context, and the bottom panel shows a trial with low-tone context.

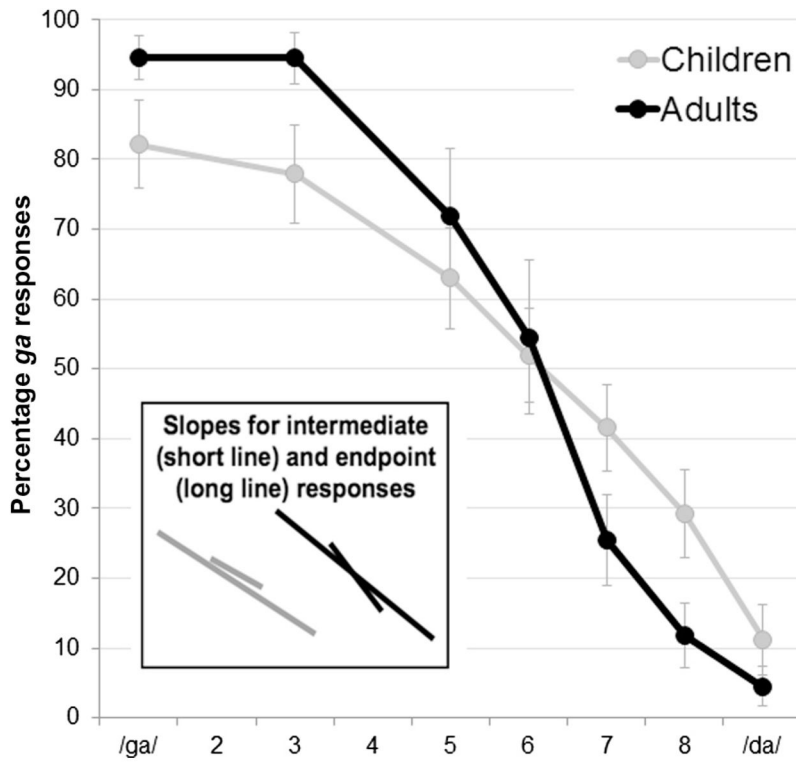


Fig. 3. Percentage *ga* responses as a function of target type. Inset: slopes for intermediate (short line) and endpoint (long line) responses, with children on the left and adults on the right.

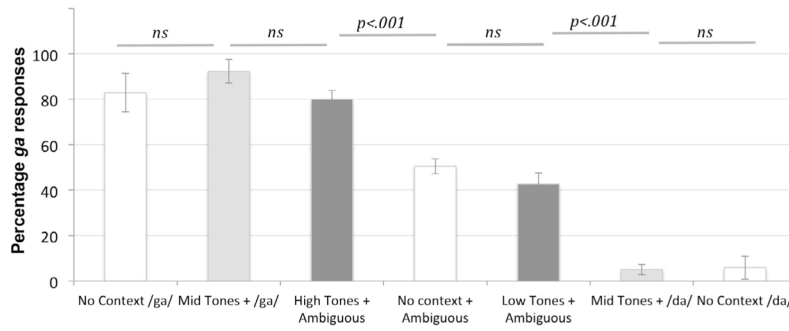


Fig. 4. Percentage *ga* responses as a function of context. Data presented are only from children who participated in both sessions. White bars represent responses from Session 1, and filled bars show responses from Session 2. Dark gray bars show responses to identical perceptually ambiguous speech targets preceded by tones sampling either higher or lower frequency spectral distributions (predicted to shift categorization toward /*ga*/ and /*da*/, respectively). Light gray bars show responses to endpoint targets preceded by tones from a neutral middle-frequency spectral distribution not expected to affect /*ga*/–/*da*/ categorization. Error bars represent standard errors. *ns*, nonsignificant.