Infants' and adults' use of temporal cues in consonant discrimination

Laurianne Cabrera

Department of Speech and Hearing Sciences

University of Washington

1417 NE 42nd St.

Seattle, WA 98105-6246

Lynne Werner

Department of Speech and Hearing Sciences

University of Washington

1417 NE 42nd St.

Seattle, WA 98105-6246

Corresponding author: Laurianne Cabrera, laurianne.cabrera@gmail.com

ABSTRACT

Objectives: Adults can use slow temporal envelope cues, or amplitude modulation (AM), to identify speech sounds in quiet. Faster AM cues and the temporal fine structure, or frequency modulation (FM), play a more important role in noise. This study assessed whether fast and slow temporal modulation cues play a similar role in infants' speech perception by comparing the ability of normal-hearing 3-month-olds and adults to use slow temporal envelope cues in discriminating consonants contrasts.

Design: English consonant-vowel syllables differing in voicing or place of articulation were processed by two tone-excited vocoders to replace the original FM cues with pure tones in 32 frequency bands. AM cues were extracted in each frequency band with two different cut-off frequencies, 256 Hz or 8 Hz. Discrimination was assessed for infants and adults using an observer-based testing method, in quiet or in a speech-shaped noise.

Results: For infants, the effect of eliminating fast AM cues was the same in quiet and in noise: a high proportion of infants discriminated when both fast and slow AM cues were available, but less than half of the infants also discriminated when only slow AM cues were preserved. For adults, the effect of eliminating fast AM cues was greater in noise than in quiet: All adults discriminated in quiet whether or not fast AM cues were available, but in noise eliminating fast AM cues reduced the percentage of adults reaching criterion from 71% to 21%.

Conclusions: In quiet, infants appear to depend on fast AM cues more than adults do. In noise, adults appear to depend on FM cues to a greater extent than infants do. However, infants and adults are similarly affected by a loss of fast AM cues in noise. Experience with the native language appears to change the relative importance of different acoustic cues for speech perception.

INTRODUCTION

Phonetic features are signaled by spectrotemporal acoustic changes in speech. Infants have been shown to distinguish fine phonetic contrasts, and this ability is shaped by exposure to a specific language (see Kuhl, 2004; Saffran *et al.*, 2006). However, how infants use spectrotemporal acoustic cues to distinguish speech sounds, and whether they use these cues like adults is still unknown. The present study explored how infants and adults use fast and slow temporal information to discriminate phonetic contrasts.

The auditory system represents the frequency components in a complex sound and changes in those components over time. Two important time scales have been identified in auditory processing (Moore, 2004): a relatively fast one, referred to as frequency modulation (FM) cues, or "temporal fine structure"; and a relatively slow one, referred to as amplitude modulation (AM) cues or "temporal envelope". FM cues represent the relatively fast fluctuations in instantaneous frequency within an auditory filter, and thus, carry information about voice pitch (*e.g.*, Zeng *et al.,* 2005 ; Xu & Pfingst, 2003). The fastest AM cues also convey information about voice pitch (*e.g.*, Kong & Zeng, 2006), as well as formant transitions, whereas the slowest AM cues convey syllabic and phonetic information (see Rosen, 1992). The auditory processing of AM cues is modeled as the operation of a central modulation filter bank (*e.g.*, Dau *et al.*, 1997a, 1997b). The auditory processing of FM cues is constrained by neural phase locking in auditory-nerve fibers, at least for slow FM rates (< 5 Hz) and low carrier frequencies (< 1 kHz). For faster FM rates and higher carrier frequencies, FM cues may be converted into AM and spectral cues as a consequence of cochlear filtering (see Moore, 2004).

The relative importance of AM and FM cues for speech perception by adults has been addressed in many studies. An early influential study by Shannon *et al.*, (1995) evaluated the role of FM cues in speech perception using noise-excited vocoders, which filter the signal into some number of frequency bands, then replace the original FM cues with noise in each band. The original speech AM cues in each frequency band were low-pass filtered with high-frequency cutoffs ranging from 16 to 500 Hz. Shannon *et al*. showed that syllable identification in quiet was poor for 1, 2, or 3 analysis bands, but was nearly perfect with 4 bands, irrespective of the cutoff frequency of the low-pass filter used to extract AM cues. This study also reported that place of articulation was an exception among phonetic features in that its identification remained poor with 4 frequency bands when FM and fast AM cues were reduced. This result indicates a greater dependence on spectral information and/or fast temporal cues for place contrasts. Further, Drullman *et al.*, (1994a, 1994b) showed that speech identification in quiet improved as the highest available AM rate increased from 4 to 16 Hz, but did not improve further with increases up to 64 Hz. Place of articulation was also reported to be more difficult to identify, especially for stop consonants. Thus, slow AM alone may be sufficient for perception of some speech features in quiet, although identification of place of articulation seems more susceptible to spectrotemporal degradation. Furthermore, when speech sounds are distorted, masked, spectrally smeared or temporally interrupted, fast AM and FM cues assume greater importance (*e.g.*, Nelson *et al.*, 2003; Gilbert *et al.*, 2007; Stone *et al.*, 2008; Ardoint & Lorenzi, 2010; Hopkins *et al.*, 2010), and the perceptual

weighting of AM and FM cues appears to change with listening conditions (*e.g.,* Fogerty, 2011; Fogerty & Humes, 2012). Thus, adults appear to require fast AM cues and FM cues to identify speech under some listening conditions.

Very few studies have investigated the development of AM and FM processing in speech, although studies with children indicate that the relative importance of different acoustic cues in speech perception continues to change into the school years (*e.g.*, Lehman & Sharf, 1989; Mayo *et al.*, 2003; Nittrouer, 2004; Nittrouer & Lowenstein, 2007). Eisenberg *et al.* (2000) showed that school-aged children (5 to 7 years) require more spectral information than adults to identify words when FM cues and fast AM cues are degraded. However, for phonetic discrimination, Bertoncini *et al.* (2009) showed no difference between 5-year-old children and adults in the use of relatively slow (<64 Hz) AM cues when spectral information was provided.

At an earlier stage of auditory and speech development, some observations suggest that fast AM cues and FM cues play an important role in infants' speech perception. Cabrera and her colleagues (Bertoncini *et al.*, 2011; Cabrera *et al.*, 2013; Cabrera *et al.*, 2015) have shown that French 6-month-old infants, like adults, are able to use AM cues to discriminate voicing (/aba/-/apa/) and place of articulation (/aba/-/ada/) contrasts in quiet, and that they can do so even when AM rates above 16 Hz are removed. However, infants required more time to habituate to speech sounds containing only AM cues below 16 Hz, suggesting that fast AM cues may be important for infants' speech perception even in quiet. Other studies have shown that infants' listening preference for "infant-directed speech" is observed only when FM and fast AM cues are preserved (Fernald & Kuhl, 1987; Spence & Freeman, 1996).

The present study directly addressed infants' and adults' use of slow AM cues in stop consonant discrimination in quiet and in noise. It extended the approach of previous studies in several respects. First, while earlier studies tested 6-month-olds, this study compared 3-month-olds and adults, with the rationale that any existing age-related change in the cues used in speech discrimination will be more evident in infants with less experience with speech. Importantly, temporal processing seems to be efficient enough to support detection of AM in nonspeech sounds at this young age (*e.g.*, Levi & Werner, 1996). Second, the slow AM cues were reduced to rates less than 8 Hz, drastically reducing voice-pitch and formant transition information. Third, rather than discriminating between exemplars of a single minimal pair (*e.g.*, /aba/-/apa/), participants discriminated between classes of consonants differing in either voicing (/pa/, /ta/, /ka/ *versus* /ba/, /da/, /ga/) or place of articulation (/pa/, /ba/ *versus* /ta/, /da/ *versus* /ka/, /ga/), allowing stronger statements to be made about feature discrimination. Finally, the observer-based procedure (Werner, 1995) was used to assess infants' discrimination performance. While the visual habituation method used in previous studies allows one to make statements on a group level, the observer-based method allows one to assess the ability of individual infants and to determine how well infants discriminate. Thus, the current study was designed to be more sensitive to age-related differences in the use of slow AM and fast AM cues in consonant discrimination than previous studies.

In quiet, both infants and adults were expected to discriminate the phonetic contrasts on the basis of fast or slow AM as observed in previous studies (Cabrera *et al.*, 2015). Nevertheless, if infants rely more on fast AM cues than adults, then eliminating fast AM cues

was expected to reduce infants' ability to discriminate the phonetic contrasts more than it does adults'. In noise, both infants and adults were expected to discriminate the phonetic contrasts when the important fast AM cues, as well as slow AM cues, were available and to have difficulty when only slow AM cues were available. Fewer participants in both age groups were expected to discriminate place of articulation than voicing when the fast AM cues were reduced and in the presence of noise (Drullman *et al*., 1994a, b; Shannon *et al* 1995; Miller & Nicely, 1955).

MATERIALS and METHODS

Participants

Participants were recruited through the Communication Studies Participant Pool at the University of Washington. Fifty-eight 3-month-old infants (10.5 weeks - 13 weeks) and 48 adults (21-30-year-olds) participated. Infants were no older than 13 weeks at the completion of testing. All infants were born full term, had no history of otitis media within 3 weeks of testing with no more than 2 prior occurrences of otitis media, had no risk factors for hearing loss, and had no history of health or developmental concerns. They also passed newborn hearing screening. At each test session, all infants were healthy and passed a tympanometric screen with a peak admittance of at least 0.2 mmhos and peak pressure between -200 and +50 daPa with a 226 Hz probe tone[1]. All adult participants reported normal hearing bilaterally and had no history of noise exposure. All adults passed a tympanometric screen with a peak admittance of at least 0.9 mmhos and peak pressure between -200 and +50 daPa with a 226 Hz probe tone. Data from an additional six 3-month-olds were excluded, because the infants were too tired, hungry, or fussy to complete the task; and data from one 3-month-old were excluded because the infant did not pass the tympanometric screen.

Stimuli

In order to assess a more general ability to discriminate voicing and place of articulation categories, several syllables contrasting in each of those features were used in the present experiment. Speech signals were recorded in a sound-attenuated room and digitized with 16-bit resolution at a 44.1-kHz sampling rate. A female English native speaker (F0=196 Hz) who was instructed to "speak clearly" produced sequences of six CVs with the vowel /a/: /ba/, /pa/, /da/, /ta/, /ka/, /ga/. Four tokens per CV were selected for their clarity. All tokens were comparable in duration, intensity, and F0 (see **Table 1**). All stimuli were equated in global root-mean-square (RMS) level.

---

[1] It is generally accepted that tympanometry with a 226 Hz probe tone is less effective in identifying middle ear effusion in young infants than is tympanometry with a 1000 Hz probe tone (e.g., Hoffman et al., 2013). However, the use of a less-than-optimal tympanometric screen cannot explain the results of this study. If we had missed infants with middle ear effusion, one would expect that their performance would be negatively affected. However, we find that infants are at least as successful as adults in reaching testing criterion in all but one condition. The 3-month-old infant who was excluded on the basis of failed tympanometry did not reach criterion in any testing phase. The inclusion of this data point, in the noise condition, would yield a success rate of 92.6% in the AM<256Hz phase (25 of 27 reaching criterion) and 44% in AM<8Hz phase (12 of 27 reaching criterion). Thus, the observed effects of age and AM condition would not change.

CV	mean duration (ms)	SD	mean pitch (Hz)	SD
ba	414.5	18.5	197.4	2.0
da	416.3	16.6	200.9	0.3
ga	410.4	17.3	197.7	1.7
pa	413.3	16.2	209.5	1.0
ta	416.7	16.4	204.3	1.0
ka	422.4	15.7	203.3	2.3

Table 1. Mean duration (ms), mean F0 (Hz) and standard deviation (SD) of 4 syllables in each Consonant-Vowel category.

The original stimuli were processed by vocoders to alter the temporal modulation rates. Tone-excited vocoders were used instead of noise-excited vocoders, because they distort speech AM cues less (*e.g.*, Kates, 2011). Two different vocoder conditions were designed. In each condition, the original speech signal was passed through a bank of 32 2nd-order gammatone filters (Patterson, 1987; Gnansia *et al*, 2009), each 1-equivalent rectangular bandwidth (ERB) wide with center frequencies (CFs) uniformly spaced along an ERB scale ranging from 80 to 8,020 Hz. The Hilbert transform was then applied to each bandpass filtered speech signal to extract the AM component and FM carrier. The FM carrier in each frequency band was replaced by a sine wave carrier with frequency at the CF of the gammatone filter and random starting phase. The AM component was low-pass filtered using a zero-phase Butterworth filter (36 dB/octave rolloff) with a cutoff frequency set to either 256 Hz (AM<256Hz) or 8 Hz (AM<8Hz). Each tone carrier was multiplied by the corresponding filtered AM function. The narrow-band speech signals were finally added up and the level of the wideband speech signal was adjusted to have the same RMS value as the input signal. Two steady speech-shaped noises were designed to have the same long-term spectrum as the syllables processed in the AM<256Hz condition and the AM< 8Hz condition, respectively.

Thus, in both conditions the resulting vocoded speech signal discarded the original (within-channel) FM speech cues but retained either the fast and slow AM speech cues (AM<256Hz), or only the slow AM speech cues (AM<8Hz), in 32 bands. Syllabic information would be preserved in both AM rate conditions; voice-pitch and formant transition information were preserved for AM<256Hz, but drastically reduced for AM<8 Hz. **Figure 1** represents the spectrograms and waveforms of one exemplar of a /pa/ syllable in each condition. We refer to the temporal cues in the AM<256Hz condition as "fast", but note that the signals in that condition contain both fast and slow AM.
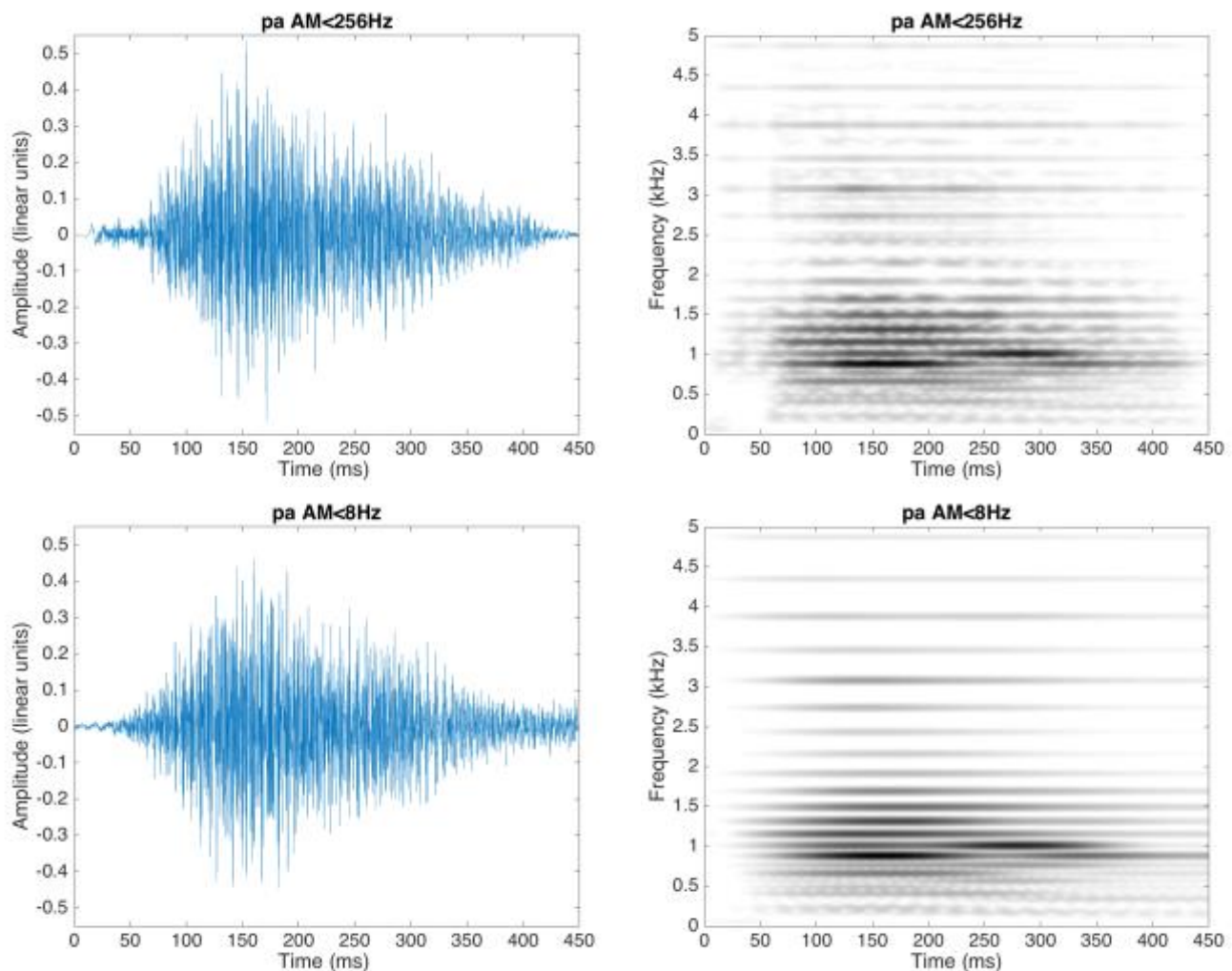
Figure 1. Spectrograms and waveforms of one exemplar of the CV category /pa/ filtered in the condition AM<256Hz (upper line) and the condition AM<8Hz (bottom line)

Material and apparatus

Infants were tested using an observer-based psychophysical procedure (Werner, 1995). During testing, infants sat on a caregiver's lap with an assistant inside a sound-attenuating booth. On the participant's right, a TV screen and two mechanical toys were placed in a Plexiglas box. The infant listened to sounds through an insert earphone (ER-2), calibrated to deliver the sounds at 65 dB SPL. The caregiver listened to music during the whole experiment and was instructed to avoid distracting the infant. The assistant listened to the experimenter's instructions and manipulated toys to keep infants facing midline. The experimenter sat outside the booth and observed through a window or monitored the output of a camera inside the booth. A microphone inside the booth enabled the experimenter to listen to the infant and assistant, and a microphone outside the booth allowed the experimenter to communicate with the assistant. A computer controlled the experiment. None of the adults involved could hear the stimuli presented to the infant.

Adults were tested in the same setup, except that they sat alone in the booth. An advantage of the observer-based procedure over procedures previously used to assess infants' discrimination of vocoded speech is that adults can be tested in the same procedure as a basis of comparison.

Procedure

The goal was to determine whether the participant detected a change in phonetic category based on voicing or place of articulation in quiet or in noise. Very few studies have assessed infants' phonetic discrimination in noise, but Nozza *et al*., (1990) found that 7-11-month-old infants showed 50%-correct detection of a change in place of articulation (/ga/ vs /ba/) at a signal to noise ratio (SNR) of -8 dB. Moreover, a pilot study using the present experimental procedure showed that both adults and 3-month-old infants were able to detect consonant changes at a -5 dB SNR. Thus, in the current noise conditions, the syllables were played in speech-shaped noise at a SNR of -5 dB.

The participant heard repeated, random exemplars from one "background" category and learned to respond when an exemplar from the "target" category was presented. Five conditions were included: voiceless (with voiced background/voiceless targets), voiced (with voiceless background/voiced targets), labial (with coronal-velar background/ labial targets) coronal (with labial-velar background/coronal targets) and velar (with labial-coronal background/velar targets). **Table 2** describes the background and target syllables in the five phonetic conditions. Each participant was tested in both AM rate conditions if possible, but in one phonetic target condition and in either quiet or noise.

Phonetic Condition	Background syllables	Target syllables
Voiceless	/ba/, /da/, /ga/	/pa/ or /ta/ or /ka/
Voiced	/pa/, /ta/, /ka/	/ba/ or /da/ or /ga/
Labial	/ta/, /da/, /ka/, /ga/	/ba/ or /pa/
Coronal	/ka/, /ga/, /pa/, /ba/	/da/ or /ta/
Velar	/ba/, /pa/, /ta/, /da/	/ga/ or /ka/

Table 2. Five phonetic conditions were designed. In each condition, background syllables were played repeatedly and randomly. When a change trial occurred, one single "target" syllable was played (randomly selected) instead of a background syllable. When a no-change trial occurred, one background syllable was played. The target syllables corresponded to the phonetic condition tested.

The experimenter initiated a trial when the participant was quiet and facing midline. There were two trial types, which occurred with equal probability during testing. On change trials a target syllable was presented once, while on no-change trials a syllable from the background category was presented. On each trial, the experimenter, blind to trial type, decided within 4 s of trial onset whether a change or no-change trial had occurred, based only

on the participant's behavior. For infants, the behavior used by experimenters to make judgments varied from infant to infant. Eye movements, increases and decreases in body movement, and facial expressions like widening of the eyes were common behaviors observed. Adults were instructed to raise their hand when they detected a change in the sounds. Computer feedback was provided to the experimenter at the end of a trial. Participants' responses were reinforced with the presentation of a mechanical toy or video for 4 s only if the experimenter correctly identified a change trial.

The experiment consisted of 3 phases, a demonstration phase and 2 test phases. The phases were presented in a fixed sequence: Participants were required to reach criterion on one phase before moving to the next. In the demonstration phase and in the first test phase the stimuli were from the AM<256Hz condition. In the second test phase the stimuli were from the AM<8Hz condition.

The purpose of the demonstration phase was to familiarize the participant with the association between the reinforcer (i.e., mechanical toy or video) and the target sounds. The probability of a change trial was 0.80, and the reinforcer was activated after every change trial regardless of the experimenter's response. The experimenter had to respond correctly based on infants' behavior on 1 change trial and 1 no-change trial within 12 trials to complete the demonstration phase and progress to the test phases.

In the following test phases, the probability of change and no-change trials was 0.5, and the experimenter was required to respond correctly on 4 of the last 5 change and 4 of the last 5 no-change trials to reach criterion. This criterion corresponds to a hit rate of 80% and a false alarm rate of 20% on the last 5 consecutive change and no-change trials, respectively. The reinforcer was activated only when the experimenter correctly identified a change trial. In the test phases the participant learned that an observable response to the phonetic change was required to activate the reinforcer. If the criterion was not reached within a maximum number of trials, the session ended and a new session was started after a short break. If the participants could not reach the criterion within a maximum number of sessions, the participant was judged to be unable to complete the phase. In the AM<256 Hz test phase the maximum number of trials was 40 and the maximum number of sessions was 4; in the AM<8 Hz test phase the maximum number of trials was 32 and maximum number of sessions was 3.

Based on previous studies, it was expected that infants would have a relatively easy time with all discriminations in the AM<256Hz condition, but a more difficult time with the AM<8 Hz condition. To accommodate the anticipated difficulty in the AM<8 Hz condition a reminder procedure, similar to the one used by Clarkson and Clifton (1995), was used to assess whether an infant's failure was due to factors such as sleepiness or boredom rather than an inability to discriminate. If a participant responded incorrectly on three consecutive trials in the AM<8 Hz test phase—responding to no-change trials or not responding to change trials in the test phase—stimuli were presented from the previously completed AM<256Hz condition. Up to 10 trials of such "reminder" trials were presented, and if the participant responded correctly on three of four consecutive trials, the participant returned to the AM<8Hz phase. If this criterion was not met, the session was discontinued, and infants were given a break or returned on another day. If a participant reached criterion in the AM<256Hz and reached criterion in three reminder periods without reaching criterion in the AM<8Hz

condition in three sessions, the participant was judged to be unable to discriminate the phonetic contrast based on the slow AM cues. Thus, when testing in the AM<8Hz condition ended, the infant had either reached criterion in the AM<8Hz condition or had failed to reach criterion in the AM<8Hz condition while still reaching criterion on the AM<256Hz reminder trials. Because the infant could still perform the discrimination in the AM<256Hz condition, we conclude that the infant's failure in the AM<8Hz condition did not result from fatigue or loss of interest.

Testing was completed in 60-min visits on 2 separate days within a 2-week period for the infants and in one visit for the adults.

To summarize: the independent variables were age, noise condition (quiet or noise), target contrast (voicing or place of articulation) and AM cutoff frequency (8 or 256 Hz). Age, noise and contrast were between-subject variables; AM cutoff frequency was varied within subject. Target feature (voiced or voiceless for the voicing contrast; labial, coronal or velar for the place contrast) was counterbalanced across subjects in each noise condition. Thirty-two infants (13 voicing, 19 place) and 20 adults (8 voicing, 12 place) were tested in quiet; 26 infants (10 voicing, 16 place) and 28 adults (12 voicing, 16 place) were tested in noise. The number of participants tested in each condition varied, because the goal was to obtain at least 8-10 participants reaching discrimination criteria in each condition. Because success rate varied across conditions, more participants were required in some conditions than in others. Two dependent variables were analyzed, the proportion of participants reaching criterion and the number of trials required to reach criterion. These variables are often used as measures of processing difficulty in infant studies (e.g., Clarkson *et al.*, 1988; Clarkson & Clifton, 1995; Lau & Werner, 2012).

RESULTS

Two analyses were conducted. First, the proportion of participants who reached criterion in the AM<256Hz and AM<8Hz conditions was compared across age group, phonetic conditions, and noise conditions. Finally, the relative difficulty of the discrimination in the AM<256Hz and AM<8Hz conditions was assessed by comparing the average number of trials required to meet criterion across phonetic and noise conditions.

The proportion of participants reaching criterion can be considered a measure of task difficulty. **Figure 2** breaks down the proportion of participants reaching criterion by condition. In quiet, 26 of 32 infants tested reached criterion in the AM<256Hz condition, but only 13 of 26 infants reached criterion in the AM<8Hz condition. All 20 adults tested in quiet reached criterion in both AM<256Hz and AM<8Hz conditions. In noise, 25 of 26 infants reached criterion in the AM<256Hz condition, but only 12 of 25 infants reached criterion in the AM<8Hz condition. Only 20 of 28 adults tested in noise reached criterion in the AM<256Hz condition, and an even smaller number of adults, 6 of 20, reached criterion in the AM<8Hz condition. Thus, it appears that infants, but not adults, have difficulty discriminating consonant contrasts with only slow AM cues in quiet. In noise, both infants and adults have difficulty with only slow AM cues.
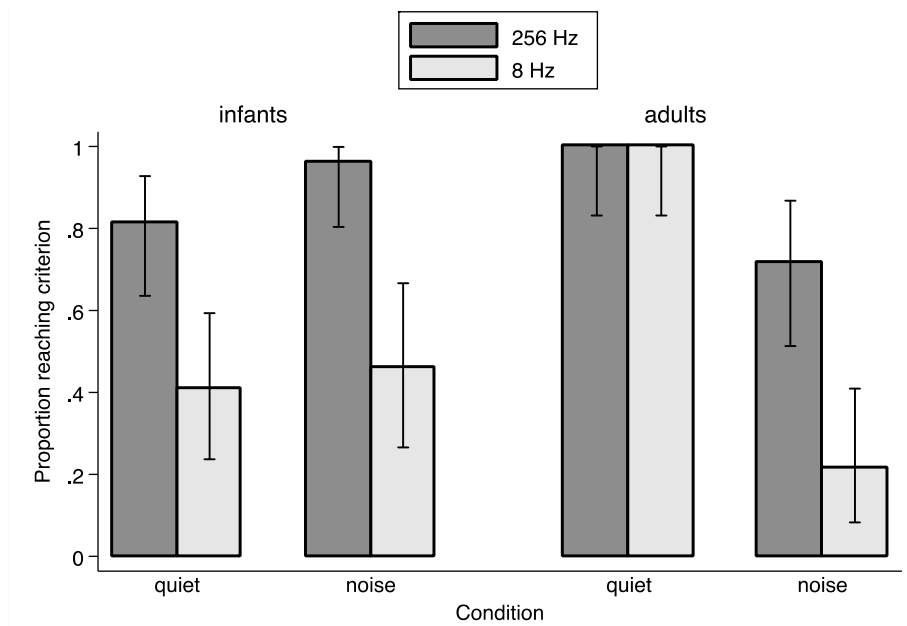
Figure 2. Percentage of infants (left) and adults (right) who reached the 80%-correct criterion in quiet and noise in the AM<256Hz (dark bars) and AM<8Hz (light bars) conditions. Error bars represent the 95% binomial confidence interval.

It was not clear whether the small number of participants reaching criterion in the AM<8Hz condition was actually greater than would be expected if the experimenter or participant simply responded randomly on each trial. To assess that possibility, a simulation of 10000 participants was run, following all constraints imposed in the actual experiment, including the probability of a change trial, maximum number of trials, the number of attempts permitted, reminder trials in the AM<8Hz conditions, and the passing criteria in each of the phases of the experiment. On each trial, a change response was assumed to occur randomly with a probability matching the rate at which change responses were recorded in all sessions run during the experiment, broken down by age and listening condition (infants/quiet 0.53; infants/noise 0.57; adults/quiet 0.44; adults/noise 0.40). Using the observed response rates takes into account the effects of response bias. The probability of an infant reaching criterion by chance in all three phases was 0.056 in quiet and 0.052 in noise; the probability of an adult reaching criterion by chance in all three phases was 0.051 in quiet and 0.045 in noise. An exact binomial test comparing the observed proportions reaching criterion with the simulation results showed that more participants reached criterion in all test phases than expected by chance (all $p < 0.005$). Thus, even though relatively few participants successfully completed the AM<8Hz condition, it is unlikely that all of them did so by guessing.

Because participants who failed in the AM<256Hz condition were not included in the AM<8Hz condition, a modified logistic regression approach, survival analysis, was used to test the effect of AM condition within age groups, noise condition, and phonetic category. The analysis compared the "survival" (i.e., the proportion reaching criterion) of participants in the AM<256Hz and AM<8Hz conditions. The survival function shows the cumulative probability that a participant who started the experiment reached criterion in each condition. The log-rank test for equality, a nonparametric statistic, was used to compare the survival functions for infants and adults, in quiet and in noise, and for voicing and place. The test

11

compares the survival functions as a whole, but provides confidence intervals for each point on the function. The functions defined by age, noise condition and phonetic category were significantly different [$X^2(7) = 29.99$, $p = 0.0001$].

The next comparisons addressed differences between age groups in quiet and in noise. The survival functions for infants and adults in quiet and in noise are shown in **Figure 3**. In quiet, the functions were significantly different for infants and adults [$X^2(1) = 11.27$, $p = 0.0008$], because fewer infants than adults reached criterion in the AM<8Hz condition. In noise, the functions were again significantly different for infants and adults [$X^2(1) = 6.14$, $p = 0.01$], but in this case fewer adults than infants reached criterion in both the AM<256Hz condition and the AM<8Hz condition. The survival functions for infants and adults are close to parallel, indicating that the effect of AM rate cutoff frequency did not differ between age groups. Thus, in quiet infants were more affected than adults by a reduction in AM information, while in noise adults seemed particularly susceptible to the elimination of FM information, but infants and adults were similarly affected by a reduction in AM information. Further analyses indicated that the effect of reducing AM information was not statistically different in quiet and in noise for infants [$X^2(1) = 0.74$, $p = 0.39$], while for adults, the effect of reducing AM information had a significantly greater effect in noise than in quiet [$X^2(1) = 19.62$, $p < 0.0001$].
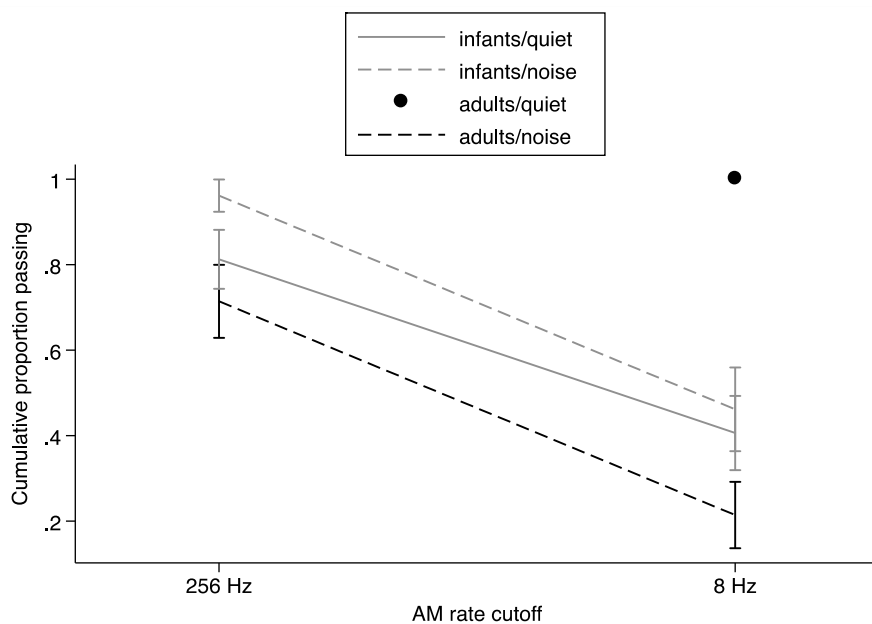

Figure 3. Survival functions, cumulative proportion of participants reaching criterion in the AM<256Hz and AM<8Hz conditions. Because all adults tested in quiet reached criterion in both conditions, only one point is plotted for that group, at 8Hz. Error bars are standard errors from the Kaplan-Meier analysis.

The next set of comparisons addressed differences across phonetic categories in quiet and noise, separately for infants and adults. The survival functions for the voicing and place of articulation contrasts are shown in **Figure 4**. For infants (top panel), it is clear that the phonetic contrast tested did not have an effect on performance. The log-rank test indicated no significant differences among the survival functions of infants [$X^2(3) = 0.88$, $p = 0.83$]. For

adults, it appears that reducing AM information had a greater effect on participants tested in noise on the place of articulation contrast, but that trend was not statistically significant [$X^2(1) = 0.42$, $p = 0.52$].
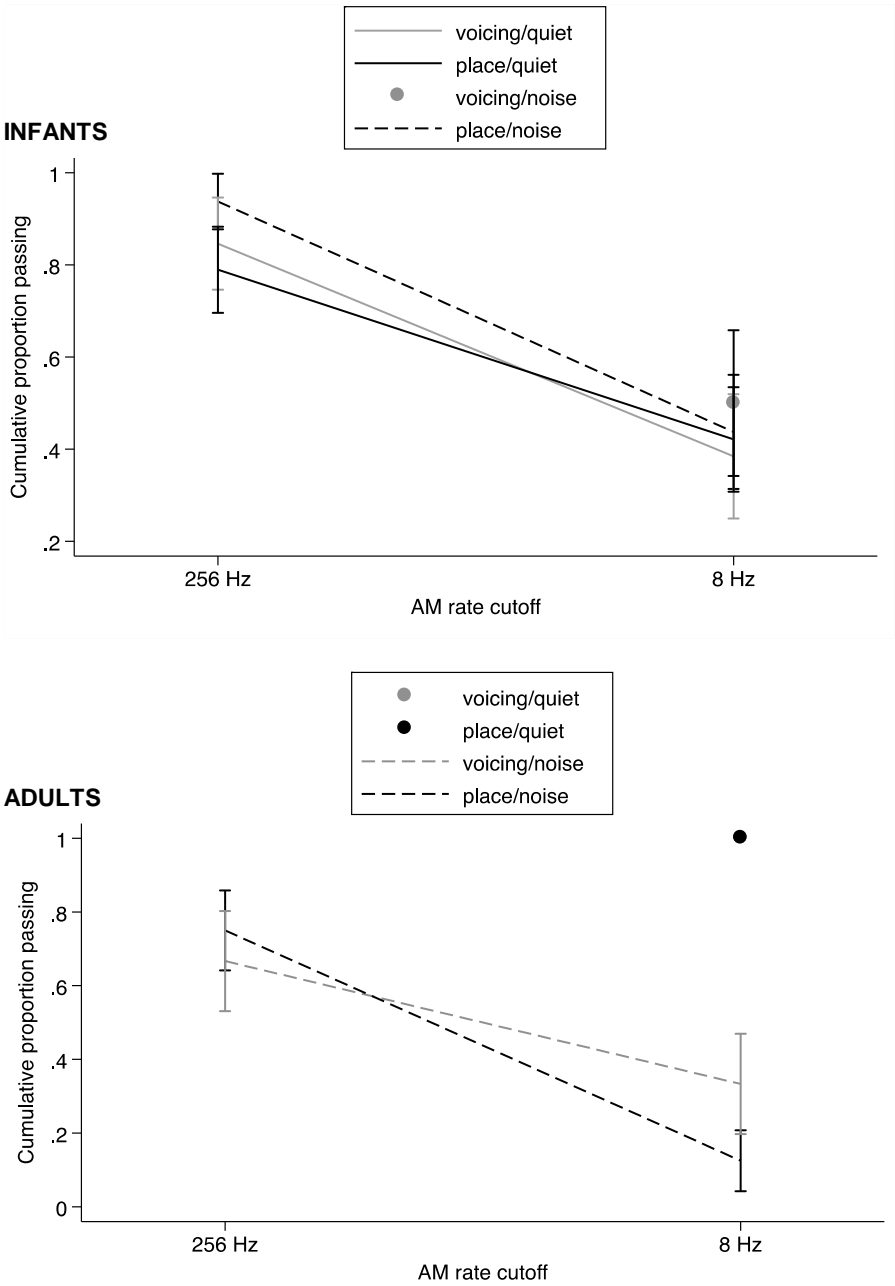

Figure 4. Survival functions, cumulative proportion of participants reaching criterion, for infants (top) and adults (bottom) for voicing and place of articulation contrasts, in quiet and in noise. A single point is plotted for infants at 8 Hz, because all infants tested on the voicing contrast in noise reached criterion at 256 Hz. A single point is plotted for adults at 8 Hz, because all adults tested on both contrasts in quiet reached criterion at 256 Hz and at 8 Hz and the symbols for the two conditions are overlaid. Error bars are standard errors from the Kaplan-Meier analysis.

A second analysis compared the number of trials required to reach criterion, another measure of relative difficulty, across age groups and noise conditions. In the following analyses, only data from participants who reached criterion in both AM<256Hz and AM<8Hz

conditions were included. However, we compared the AM<256Hz performance of infants who completed only the AM<256Hz condition to those who completed both AM conditions, to ensure that the infants failing the AM<8Hz condition were not simply poor performers. In quiet, infants who only reached criterion in the AM<256Hz condition required an average of 18 trials (SD = 9), while infants who reached criterion in both conditions required an average of 25 trials (SD = 10) to reach criterion. In noise, the corresponding means are 27 trials (SD = 9) and 24 trials (SD = 9), respectively. The mean number of trials to did not differ between the two groups of infants in quiet [t(24) = -1.78, p = 0.09] or in noise [t(23) = -0.71, p = 0.48].

In all conditions but one, infants and adults needed about 2 sessions on average to reach criterion; the exception was for adults tested in the AM<256Hz condition in quiet, in which adults required only 1.2 sessions on average. **Figure 5** shows the average number of trials in the session in which criterion was reached in the AM<256Hz and AM<8Hz conditions. Infants who reached criterion in both conditions needed fewer trials to reach criterion in the AM<8Hz (second) condition than in the AM<256Hz (first) condition. Adults showed a different pattern of results: While adults who reached criterion in the AM<8Hz condition in noise took fewer trials to reach criterion in the AM<8Hz (second) condition than in the AM<256Hz (first) condition, adults tested in quiet took more trials to reach criterion in the AM<8Hz (second) condition than in the AM<256Hz (first) condition. Three-way analyses of variance (ANOVA) were used to test the effect of AM rate, Phonetic category (Voicing *versus* Place) and Noise condition on the number of trials to criterion in each age group. For infants, the analysis showed a significant effect of AM rate [F(1, 21) = 14.23, p = 0.001] indicating that infants required fewer trials to reach criterion in the second AM<8Hz phase than in the first AM<256Hz phase. No other effect was statistically significant [Noise: F(1,22) = 0.853, p=.366; Phonetic: F(1,21) = 1.87, p = 0.186; Noise × Phonetic: F(1,21) = 3.14, p = 0.091; AM rate × Noise: F(1,21) = 0.006, p = 0.94; AM rate × Phonetic, F(1,21) = 0.046, p = 0.833; AM rate × Phonetic × Noise, F(1,21) = 0.046, p = 0.833] .

For adults, the ANOVA revealed no effect of AM rate [F(1,22) = 0.87, p = 0.77], Phonetic contrast [F(1,22) = 0.46, p =. 0505] or noise [F(1,22) = 1.24, p =. 0287]. The analysis of the interactions revealed a marginal interaction between AM rate and Noise [F(1, 22) = 3.99, p = 0.058], but no interactions involving Phonetic contrast [Phonetic × Noise: F(1,22) = 1.24, p = 0.277; AM rate × Phonetic: F(1,22) = 0.44, p = 0.512, AM rate × Noise × Phonetic: F(1,22) = 0.365, p = 0.55]. Although the interaction between AM rate and Noise was only marginally significant, we conducted follow-up analyses because the original hypothesis was that this interaction would be significant. Pairwise comparisons indicated that adults required more trials to reach criterion in the second AM<8Hz phase in quiet than in the first AM<256Hz phase (p=.08), but no difference was observed in noise. Thus, there is a suggestion in the data that adults who completed both AM conditions were able to apply what they learned in the AM<256Hz condition to learning the AM<8Hz condition when tested in noise, but not in quiet. Contrary to what we expected, the phonetic category did not affect the number of trials to criterion for adults.
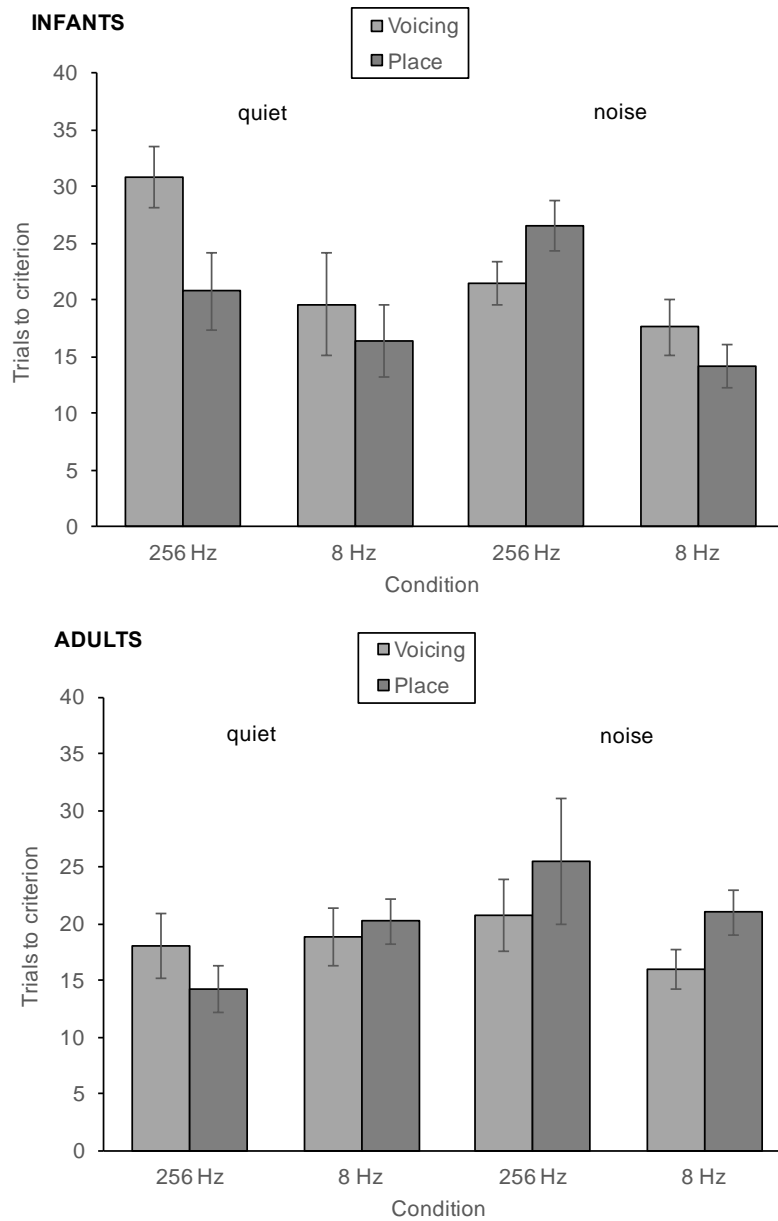
Figure 5. Mean number of trials to reach criterion for infants (top panel) and adults (bottom panel) in AM<256Hz and AM<8Hz conditions in quiet (left) and in noise (right) for voicing target (light bars) and place targets (dark bars). Error bars represent standard error.

DISCUSSION

The present study explored infants' and adults' reliance on slow AM cues in discrimination of stop consonants varying in voicing or place of articulation. The results showed that 3-month-old infants are able to use slow temporal cues to discriminate between phonetic categories differing in those features: More infants than expected by chance were able to discriminate between phonetic categories when fast AM cues were reduced. These findings replicated previous experiments assessing 6-month-olds' discrimination of voicing and place of articulation contrasts on the basis of AM cues (Bertoncini *et al.*, 2011; Cabrera *et al.*, 2013, 2015). Furthermore, the present experiment extended the previous results by using multiple syllables in each phonetic category, rather than a single consonant. Infants were able to detect a change between phonetic categories using AM cues.

The results for adults are consistent with those of previous studies. Adults readily discriminate between speech sounds when FM cues are reduced, even when fast AM cues are also eliminated as long as speech is presented in quiet (*e.g.*, Shannon *et al.*, 1995; Zeng *et al.*, 2005). However, when noise is introduced adults are less likely to discriminate when fast AM cues are eliminated.

Infants and adults were quite similar in their task performance in many respects. Infants were at least as likely as adults to master the phonetic discrimination in all conditions except the condition in which only slow-AM cues were available in quiet. This result held for discrimination of both voicing and place of articulation. What that meant was that fewer infants reached criterion performance with only slow-AM cues than with both slow- and fast-AM cues in both quiet and noise. In contrast, fewer adults reached criterion performance with only slow-AM cues than with both slow- and fast-AM cues in noise, but not in quiet. One interpretation of infants' difficulty discriminating between sounds based on slow-AM cues alone in quiet is that they depend more heavily on fast-AM cues than adults do. Because fast-AM cues carry information about F0 and formant transitions, this interpretation is consistent with infants' early preference for exaggerated prosodic cues and their ability to use F0-related variations to discriminate between speech sounds (Fernald & Kuhl, 1987; Mehler *et al.*, 1988; Spence & Freeman, 1996).

There are naturally other interpretations of infants' poor performance with slow-AM cues alone. For example, because the slow AM condition was always tested after the slow+fast AM condition, it is possible that many infants were confused by the change in the stimulus. However, note that during the test phase in which infants failed to discriminate between sounds based on slow AM cues, they were still able to discriminate based on slow+fast AM on the reminder trials. That observation argues against the idea that the stimulus change generally disrupted infants' ability to perform the task. Furthermore, there are several examples in the literature of infants' successful completion of experiments involving blocks of trials in different stimulus conditions (e.g., Spetner & Olsho, 1990; Werner *et al.*, 2013).

Another possibility is that infants simply need more cues than adults do to discriminate between sounds. Perhaps their performance would have been just as disrupted by the reduction of slow AM as it was by the reduction of fast AM. That explanation cannot be

eliminated given the design of the stimuli in the present experiment[2]. Note, however, that there have been several demonstrations that infants can continue to discriminate between speech sounds when some naturally occurring cues are eliminated. For example, in the current experiment infants' discrimination in most conditions was apparently unaffected by the elimination of FM cues: When both fast and slow AM information was available, nearly all infants could perform the task. In addition, numerous studies have demonstrated that infants discriminate between synthetic speech sounds containing only a subset of the cues that distinguish between natural productions (e.g., Eimas *et al.*, 1971; Morse, 1972). Nonetheless, it remains to future research to determine whether a reduction in any AM cue could have produced the pattern of results seen for infants here.

One rather surprising result of this study was that infants were actually more likely than adults to master the phonetic discrimination when both slow AM and fast AM cues, but no FM cues, were available in noise. This finding suggests that in the absence of FM cues, nearly all infants, but only about 71% of adults, were able to make use of fast AM to detect changes in noise, a condition in which the slow AM would have been degraded. Why some adults were unable to take advantage of fast AM cues in noise is not clear, but it is possible that adults would have been more flexible in their approach had they known that the sounds they were hearing were speech. Previous research has demonstrated that knowing that the signal is speech improves the ability to identify vocoded speech (e.g., Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008). On debriefing, only a few adult participants in the current study reported that they thought the sounds were speech, and none of them could identify the consonants involved. Because we attempted to match conditions for infants and adults as far as possible here, the adults were not informed of the nature of the stimuli because infants cannot be so informed, at last not directly. Future research might address this issue by comparing adults performance with and without that information or by developing a method to inform both infants and adults nonverbally.

It was expected that the detection of a change in place of articulation would be more susceptible to the reduction of fast temporal information and to the effects of background noise than would detection of a change in voicing (*e.g.*, Başkent, 2006; Miller & Nicely, 1955; Shannon *et al.*, 1995). However, no difference between voicing and place of articulation was observed for adults in quiet here. Moreover, voicing and place of articulation discrimination was similarly disrupted by the elimination of fast AM cues in noise. Although, detection of voicing in quiet was slightly more difficult for infants than place when FM cues were reduced, no difference was observed between phonetic contrasts in noise. Thus, the present experiment does not show that the discrimination of place of articulation is more difficult in noise or when the fast temporal cues were reduced for either infants or adults. It is possible that learning to detect one place of articulation category only (i.e., either labial, coronal or velar consonants) requires less fine spectral and temporal details than identification

---

[2] It is important to note that studies in adults show that the slowest AM cues are recovered (i.e., reconstructed) at the output of the cochlear filters in the normal ear based on higher-rate acoustic envelope cues. Swaminathan and Heinz (2012) evaluated the neural coding of vocoded speech from a physiologically based auditory nerve model and showed that low frequency AM is likely reintroduced in the auditory system after filtering. In other words, filtering out the low frequency AM does not affect performance, does not eliminate fast AM information. In the present paper, we choose to compare AM<256Hz and AM<8Hz conditions in order to avoid any reconstruction of the acoustic envelope at the output of the auditory nerve filters.

of consonants based on an open-set of possibilities (i.e., set of 16 possibilities as in Miller & Nicely, 1955; Shannon *et al.*, 1995).

Finally, there are several potential explanations for infants' apparent reliance on fast AM cues in discriminating between phonetic categories of vocoded speech in quiet. It is unlikely that this developmental difference is related to an immaturity of temporal resolution, that is, the representation of AM in the auditory system. Although temporal processing continues to mature in childhood (*e.g.*, Buss *et al.*, 1999), 3-month-olds have been shown to detect AM in non-speech sounds even at fast AM rates (*e.g.*, Levi & Werner, 1996). Moreover, one would predict that immature temporal resolution would result in a reduction in the ability to use fast AM. Furthermore, 3-month-olds have mature spectral resolution for frequencies below 4000 Hz (Spetner & Olsho, 1990; Folsom & Wynne, 1987), so it is unlikely that immature spectral resolution forces them to rely on other cues. Another possible explanation is that infants have access to speech cues, but do not weight speech temporal information as adults do. Developmental changes in speech cue weighting have been described in older children (*e.g.*, Nittrouer & Lowenstein, 2007; Nittrouer *et al.,* 2009) and it is well established that infants become "attuned" to their native language over the course of the first year of life (see Kuhl, 2004; Werker & Tees, 1984). That attunement may be related to the discovery of the most informative acoustic cues in native speech (see Cabrera *et al.*, 2015). Thus, it is possible that the reliance on fast speech AM cues changes with greater exposure to speech sounds. More studies comparing the specific role of slow and fast temporal cues of speech during development are needed to explore such hypotheses.

An interesting question is why some infants, but not others, were able to learn to use slow AM cues. One hypothesis is that some infants may have actually learned to use the slow AM cues in the course of the experiment. If an infant quickly learned the correlation between fast and slow cues in the initial test, that infant would then be able to use that knowledge, or generalize, in the slow AM condition. Increasing exposure to the combined fast and slow cues might enable a greater number of infants to use slow AM when it was the only cue available.

CONCLUSIONS

The present study compared how 3-month-olds and adults use slow (<8 Hz) AM cues to detect a change in voicing or place of articulation in initial stop consonants in quiet and in noise. Results indicated that both infants and adults are able to use slow AM cues to detect a change in phonetic category. Adults, but not infants, had difficulty detecting phonetic changes in noise when both fast and slow AM cues were available. However, while adults easily detected phonetic changes in quiet when only slow AM cues were available, only about half of the infants could do so. The results may indicate that infants rely more on fast AM cues, related to F0 and formant transition information, in speech discrimination, even in quiet. Thus, infants and adults may not use speech temporal information in the same way. Exposure to the native language may drive the development of adultlike weighting strategies in speech perception.

REFERENCES

Ardoint, M., & Lorenzi, C. (2010). Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hearing Research*, *260*(1-2), 89–95.

Başkent, D. (2006). Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *The Journal of the Acoustical Society of America*, *120*(5 Pt 1), 2908–2925.

Bertoncini, J., Nazzi, T., Cabrera, L., & Lorenzi, C. (2011). Six-month-old infants discriminate voicing on the basis of temporal envelope cues (L). *The Journal of the Acoustical Society of America*, *129*(5), 2761-2764.

Bertoncini, J., Serniclaes, W., & Lorenzi, C. (2009). Discrimination of speech sounds based upon temporal envelope versus fine structure cues in 5-to 7-year-old children. *Journal of Speech, Language, and Hearing Research, 52(3)*, 682–695.

Buss, E., Hall, J. W., 3rd, Grose, J. H., & Dev, M. B. (1999). Development of adult-like performance in backward, simultaneous, and forward masking. *Journal of Speech, Language, and Hearing Research*, *42*(4), 844–849.

Cabrera, L., Bertoncini, J., & Lorenzi, C. (2013). Perception of Speech Modulation Cues by 6-Month-Old Infants. *Journal of Speech, Language, and Hearing Research, 56,* 1733–1744.

Cabrera, L., Lorenzi, C., and Bertoncini, J. (2015). Infants discriminate voicing and place of articulation with reduced spectral and temporal modulation cues *Journal of Speech, Language, and Hearing Research, 58*, 1033–1042.

Cabrera, L., Tsao, F. M., Liu, H. M., Li, L. Y., Hu, Y. H., Lorenzi, C., & Bertoncini, J. (2015). The perception of speech modulation cues in lexical tones is guided by early language-specific experience. *Frontiers in psychology*, *6*.

Clarkson, M. G., & Clifton, R. K. (1995). Infants' pitch perception: Inharmonic tonal complexes. *The Journal of the Acoustical Society of America*, *98*(3), 1372-1379.

Clarkson, M. G., Clifton, R. K., & Perris, E. E. (1988). Infant timbre perception: Discrimination of spectral envelopes. *Perception & psychophysics*, *43*(1), 15-20.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, *102*(5 Pt 1), 2892–2905.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *The Journal of the Acoustical Society of America*, *102*(5 Pt 1), 2906–2919.

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*, 2670.

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, *95*, 1053.

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

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spec- tral cues as a function of age. The Journal of the Acoustical Society of America, 107(5, Pt. 1), 2704–2710.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293.

Fogerty, D. (2011). Perceptual weighting of the envelope and fine structure across frequency bands for sentence intelligibility: effect of interruption at the syllabic-rate and periodic-rate of speech. *The Journal of the Acoustical Society of America*, *130*(1), 489–500. http://doi.org/10.1121/1.3592220

Fogerty, D., & Humes, L. E. (2012). A correlational method to concurrently measure envelope and temporal fine structure weights: effects of age, cochlear pathology, and spectral shaping. *The Journal of the Acoustical Society of America*, *132*(3), 1679–1689.

Folsom, R. C., & Wynne, M. K. (1987). Auditory brain stem responses from human adults and infants: Wave v tuning curves. *Journal of the Acoustical Society of America, 81*, 412-417.

Gilbert, G., Bergeras, I., Voillery, D., & Lorenzi, C. (2007). Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. *The Journal of the Acoustical Society of America*, *122*(3), 1336–1339. http://doi.org/10.1121/1.2756161

Gnansia, D., Péan, V., Meyer, B., & Lorenzi, C. (2009). Effects of spectral smearing and temporal fine structure degradation on speech masking release. *The Journal of the Acoustical Society of America*, *125*(6), 4023–4033.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 460.

Hoffmann, A., Deuster, D., Rosslau, K., Knief, A., Zehnhoff-Dinnesen, A. A., & Schmidt, C. M. (2013). Feasibility of 1000 Hz tympanometry in infants: Tympanometric trace classification and choice of probe tone in relation to age. *International Journal of Pediatric Otorhinolaryngology, 77*, 1198-1203.

Hopkins, K., Moore, B. C. J., & Stone, M. A. (2010). The effects of the addition of low-level, low-noise noise on the intelligibility of sentences processed to remove temporal envelope information. *The Journal of the Acoustical Society of America*, *128*(4), 2150–2161. http://doi.org/10.1121/1.3478773

Kates, J. M. (2011). Spectro-temporal envelope changes caused by temporal fine structure modification. *The Journal of the Acoustical Society of America*, *129*(6), 3981–3990.

Kong, Y.-Y., and Zeng, F.-G. (2006). "Temporal and spectral cues in Mandarin tone recognition," *The Journal of the Acoustical Society of America*, *120*, 2830–2840.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–843.

Lau, B. K., & Werner, L. A. (2012). Perception of missing fundamental pitch by 3-and 4-month-old human infants. *The Journal of the Acoustical Society of America*, *132*(6), 3874-3882.

Lehman, M. E., & Sharf, D. J. (1989). Perception/production relationships in the development of the vowel duration cue to final consonant voicing. *Journal of Speech, Language, and Hearing Research*, *32*(4), 803–815.

Levi, E. C., & Werner, L. A. (1996). Amplitude modulation detection in infancy: Update on 3-month-olds. *Association for Research in Otolaryngology*, *19*, 142–150.

Mayo, C., Scobbie, J. M., Hewlett, N., & Waters, D. (2003). The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *Journal of Speech, Language, and Hearing Research*, *46*(5), 1184–1196.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*, 338–352.

Moore, B. C. (2004). *An introduction to the psychology of hearing* (Vol. 4). Academic press San Diego.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology, 14*, 447-492.

Nelson, P. B., Jin, S.-H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, *113*(2), 961–968.

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America*, *115*(4), 1777.

Nittrouer, S., & Lowenstein, J. H. (2007). Children's weighting strategies for word-final stop voicing are not explained by auditory sensitivities. *Journal of Speech, Language and Hearing Research*, *50*(1), 58.

Nittrouer, S., Lowenstein, J. H., & Packer, R. R. (2009). Children discover the spectral skeletons in their native language before the amplitude envelopes. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(4), 1245–1253.

Nozza, R. J., Rossman, R. N., Bond, L. C., & Miller, S. L. (1990). Infant speech-sound discrimination in noise. *The Journal of the Acoustical Society of America*, *87*(1), 339-350.

Patterson, R. D. (1987). A pulse ribbon model of monaural phase perception. *The Journal of the Acoustical Society of America*, *82*(5), 1560–1586.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In *Handbook of child psychology* (D. Kuhn & R. Siegler, Vol. 2, pp. 58–108).

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (New York, N.Y.)*, *270*(5234), 303–304.

Spence, M. J., & Freeman, M. S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behavior and Development*, *19*(2), 199–212.

Spetner, N. B., & Olsho, L. W. (1990). Auditory frequency resolution in human infancy. *Child Development, 61*, 632-652.

Stone, M. A., Füllgrabe, C., & Moore, B. C. (2008). Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region. *The Journal of the Acoustical Society of America*, *124*, 2272–2282.

Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *The Journal of Neuroscience*, *32*(5), 1747-1756.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.

Werner, L. A. (2013). Infants' detection and discrimination of sounds in modulated maskers. *Journal of the Acoustical Society of America, 133*, 4156-4167.

Xu, L., and Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception. *The Journal of the Acoustical Society of America 114*, 3024–3027.

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., … Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2293–2298.