Modified spectral tilt affects infants’
native-language discrimination
of approximants and vowels

Elizabeth Francis Beach
National Acoustic Laboratories, 16 University Avenue, Macquarie University,
Sydney, New South Wales, 2109, Australia
elizabeth.beach@nal.gov.au

William Noble
School of Behavioural, Cognitive and Social Science, University of New England,
Armidale, New South Wales, 2350, Australia

Christine Kitamura
MARCS Institute for Brain, Behaviour and Development, University of Western Sydney,
Penrith, New South Wales, 2750, Australia

Abstract: This study’s aim was to determine if 6- and 9-month-old infants discriminate approximants and vowels when the spectral shape is modified to emphasize high- or low-frequency information. Infants were presented with /r/-/l/ and /æ/-/ in three conditions: (a) unmodified; (b) −6 dB/octave tilt; and (c) +6 dB/octave tilt. Six-month-olds discriminated /r/-/l/ in conditions (a) and (b), and /æ/-/ in conditions (a) and (c), but 9-month-olds only discriminated when unmodified. The results reflect native-language attunement. Six-month-olds discriminate spectrally modified sounds that emphasize relevant cues, but by 9 months, infants are sensitive to the native spectral profiles of speech.

1. Introduction
The trajectory of infant speech perception follows a well-known path from an early phase in which very young infants rely on the acoustic properties of speech to perceive the differences between native and non-native contrasts. Between 6 and 9 months of age perceptual reorganization occurs, and there is a decline in the discrimination of contrasts that are non-phonemic in the native language, vowels first around 6 months (Kuhl et al., 1992; Polka and Werker, 1994) and consonants around 9 months (Kuhl et al., 2008). Thus, by 9 months of age speech perception has become linguistically driven and is constrained toward language-specific phonemic contrasts. This account of infant language development is supported by numerous studies which show that young infants can discriminate a wide range of native and non-native speech contrasts (Eimas et al., 1971; Trehub, 1976; Pegg and Werker, 1997), so long as they are sufficiently acoustically salient (see Narayan et al., 2010), but by the age of around 9 months, infants only discriminate native speech contrasts.

More recently attention has turned to infants’ perception of spectrally modified speech. Cabrera et al. (2013) and Cabrera et al. (2015) showed that 6-month-olds can discriminate vocoded consonants, /b/-/p/ and /b/-/d/, in which the number of channels varied between 4 and 32 and the spectral and temporal information available was systematically degraded. Six-month-olds can also discriminate similarly degraded vowels, /l/-/l/, although they were less able to cope with degraded spectral information when compared to older children and adults (Warner-Czyz et al., 2014).

In a study investigating infants’ perception of spectrally modified fricatives, Beach and Kitamura (2011) found new evidence of native-language attunement between the ages of 6 and 9 months. When the natural speech spectrum of the fricatives, /ʃ/-/ʃ/, was modified to emphasize either low- or high-frequency information, 6-month-olds could discriminate the contrast whether spectrally modified or not, but 9-month-olds only discriminated the unmodified contrast. Initially motivated by the
question of whether a modified spectral tilt might be a beneficial amplification strategy for improving speech perception for infants with hearing loss, we suggested that because spectral tilt interferes with native speech perception, and assuming this also occurs in other speech contrasts, this should be taken into account when deciding whether to modify the spectral shape of speech amplified in hearing aids. We argued that because older infants’ speech perception is more constrained than that of younger infants, older infants may have more specific amplification requirements than younger infants, who are likely to adapt more readily to spectrally tilted amplified speech.

In this study, we return to the work reported by Beach and Kitamura (2011) to examine whether the developmental effect found for fricative discrimination is indicative of a generalized effect that applies to a range of speech contrasts, or whether it is peculiar to consonants whose energy is concentrated in the high-frequency region. Here the findings of Beach and Kitamura (2011) are extended to two new classes of sounds, approximants and vowels, characterized by a concentration of energy in the mid- and low-frequency range, respectively.

The approximants used in the study are /r/ and /l/. Although classed as consonants, approximants are vowel-like and exhibit characteristic formant patterns (Harrington and Cassidy, 1999). The most notable difference between /r/ and /l/ is in the frequency value of the third formant (F3). Whereas /r/ is characterized by a low F3 (between 1300 and 1800 Hz for males, higher for females), and a corresponding steep transition to the adjacent vowel, the F3 of /l/ is higher, and often difficult to identify due to the attenuating effect of an anti-resonance that occurs around the same frequency (Harrington and Cassidy, 1999). The vowels in the study are short monophthongs from Australian English: /ɔ/ (as in “hot”) and /ʌ/ (as in “hat”). Adjacent to one another in the lower right portion of F1-F2 vowel space, /ɔ/ is a rounded back vowel; unrounded /ʌ/ is lower, and not as far back.

In this study, 6- and 9-month-old infants were tested for their discrimination of /r/-/l/ and /ɔ/-/ʌ/ in one of three conditions: Unmodified normal speech; speech with a positive 6 dB/octave spectral tilt; or speech with a negative 6 dB/octave spectral tilt. Modifying the natural spectral tilt of approximants and vowels has an impact on their distinctive acoustic attributes. When positive tilt is applied, the intensity of frequency information above 1000 Hz is increased, which should facilitate discrimination of /r/-/l/ because it amplifies the F3 region around 2000–3000 Hz, where the difference between the two approximants occurs. The negative tilt, on the other hand, was expected to have the opposite effect as it reduces emphasis to the higher frequencies and may therefore obscure the differences between the approximants, making discrimination difficult. For vowels, it was expected that the negative tilt would facilitate discrimination because it adds emphasis to the low-frequency portion of the spectrum, where F1 and F2, which differentiate the vowels, are located (Delattre et al., 1951; Peterson and Barney, 1952).

Over and above these acoustic-based predictions, it was expected that the results would replicate those of Beach and Kitamura (2011). That is, 6-month-olds, who perceive speech in a language-general acoustic manner, would discriminate /r/-/l/ and /ɔ/-/ʌ/ regardless of the spectral tilt condition, while the older infants, who have progressed from a language-general to a language-specific mode of perception, would find that negative and positive spectral tilts interfere with the stimuli such that discrimination is not possible.

2. Method

Six- and nine-month-old participants were tested for their discrimination of /r/-/l/ and /ɔ/-/ʌ/ using an habituation-dishabituation procedure identical to that described in Beach and Kitamura (2011), except that approximants and vowels were used as stimuli instead of fricatives. For each contrast, infants were assigned to one of three conditions: (a) normal speech, presented unmodified; (b) negative tilt, with –6 dB/octave spectral tilt; or (c) positive tilt, with +6 dB/octave spectral tilt.

2.1 Participants

A total of 192 infants participated. Half completed the discrimination task with approximants, and half completed the task with vowels. Equal numbers of 6- and 9-month-olds completed each task, as shown in Table 1. All infants were from homes in which Australian English was the primary language, and parental questionnaires confirmed that all infants had passed their newborn hearing screening test, were born full term, had no prior or current middle ear infection, and were healthy on the day of testing. Participants were reimbursed for travel costs and received a small gift.
2.2 Speech stimuli

Four tokens each of /li+/ and /ri+/ were recorded by an adult female speaker of Australian English. Four measures were taken for each token: Total duration, F0 at the vowel midpoint, frequency of F3 at the point of its onset, and F3 at the vowel midpoint (Table 2). Measures of duration, F0, and F3 at vowel midpoint were well matched amongst all tokens of /li+/ and /ri+, but the tokens differed markedly on the critical F3 onset measure.

The vowel syllables, /Ot/ and /Æt/, were recorded by the same speaker. Four exemplars of each syllable were selected as stimuli and for each token, syllable duration, F0 at vowel midpoint, and the frequencies of the first three formants at vowel midpoint were measured. Syllable duration and F0 were comparable amongst the tokens, and the formant frequency values were comparable to those reported for Australian English vowels (Butcher, 2006; Cox, 2006). The vowels were well differentiated in terms of F1 and F2, but the frequency of F3 was similar for both vowels at approximately 2900 Hz.

As in Beach and Kitamura (2011), two fast Fourier transform filters were applied to the syllables to modify their spectral tilt for use in the positive and negative tilt conditions. The negative tilt condition increased emphasis on frequencies below 1000 Hz and reduced emphasis on frequencies higher than 1000 Hz. The positive tilt condition increased emphasis on frequencies over 1000 Hz and reduced emphasis on the lower frequencies. For the normal speech condition, the original speech sample was used. Figure 1 shows the impact of spectral tilting on the long-term average speech spectra of the vowels and approximants in each condition.

2.3 Materials and apparatus

The experiments were conducted using the same sound-attenuated test rooms and equipment as for the fricative study (see Beach and Kitamura, 2011). In the test room, infants sat on their parent’s lap facing a television screen and speaker from which audio stimuli were presented. At the beginning of each trial an attention-getting stimulus was played to attract the infant’s attention to the screen. During trials the stimulus switched to a brightly colored bull’s-eye which was shown with the speech stimuli. A video camera positioned opposite the infant transmitted vision to the adjacent room where the experimenter judged and recorded the infant’s head and eye movements.

2.4 Procedure

Discrimination was tested using the infant-controlled habituation-dishabituation procedure described in Beach and Kitamura (2011). Trials began when the infant fixated the attention-getting stimulus for at least 3 s and ended when the infant looked away for more than 1.2 s, or when a total of 30 s had elapsed. First, infants were habituated to the habituation stimulus (/ri+/ for approximants, /Ot/ for vowels). The habituation stimulus was presented on repeated trials until mean fixation duration of two trials declined by at least 25%.

Table 1. Mean age, age range and number of participants in each spectral tilt condition. Infants were excluded if they failed to habituate or cried.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Approximants</th>
<th>Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean age (range)</td>
<td>Σ (data excluded)</td>
</tr>
<tr>
<td>Normal speech</td>
<td>6.0 (5.5–6.4)</td>
<td>16 (3)</td>
</tr>
<tr>
<td></td>
<td>8.9 (8.6–9.3)</td>
<td>16 (1)</td>
</tr>
<tr>
<td>Negative tilt</td>
<td>6.1 (5.7–6.5)</td>
<td>16 (2)</td>
</tr>
<tr>
<td></td>
<td>8.8 (8.6–9.3)</td>
<td>16 (5)</td>
</tr>
<tr>
<td>Positive tilt</td>
<td>6.0 (5.6–6.5)</td>
<td>16 (0)</td>
</tr>
<tr>
<td></td>
<td>8.9 (8.6–9.3)</td>
<td>16 (5)</td>
</tr>
</tbody>
</table>

Table 2. Mean acoustic measures averaged across four tokens of /li/ and /ri/ and four tokens of /Ot/ and /Æt/. Standard deviations are in parentheses.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Duration (msec)</th>
<th>F0 at vowel midpoint (Hz)</th>
<th>F1 at vowel midpoint (Hz)</th>
<th>F2 at vowel midpoint (Hz)</th>
<th>F3 at onset (Hz)</th>
<th>F3 at vowel midpoint (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/Ot/</td>
<td>528 (28)</td>
<td>172 (2)</td>
<td>894 (18)</td>
<td>1377 (19)</td>
<td>—</td>
<td>2926 (32)</td>
</tr>
<tr>
<td>/ri/</td>
<td>541 (33)</td>
<td>175 (9)</td>
<td>728 (22)</td>
<td>1072 (13)</td>
<td>—</td>
<td>2889 (29)</td>
</tr>
<tr>
<td>/li/</td>
<td>669 (26)</td>
<td>150 (0.5)</td>
<td>—</td>
<td>—</td>
<td>3005 (93)</td>
<td>3200 (23)</td>
</tr>
<tr>
<td>/ri/</td>
<td>645 (29)</td>
<td>149 (3)</td>
<td>—</td>
<td>—</td>
<td>2059 (55)</td>
<td>3164 (43)</td>
</tr>
</tbody>
</table>
by 50% of the mean fixation duration of the first two trials. If the habituation criterion was not met after 30 trials, the procedure was discontinued. Once the habituation criterion was met, two no-change control trials (of the habituation stimulus) were presented to ensure that infants had habituated and did not show spontaneous recovery (defined as an increase of at least 100% in mean fixation duration in control trials vs the final two habituation trials). Any infants who did not meet this criterion were deemed not to have habituated, and their data were excluded. The control trials were followed by two test trials that comprised the new test stimulus alternating with the habituation stimulus: liː riː liː riː, etc., for approximants and /æt/, /ɜt/, /ʌt/ for vowels. Infants who showed recovery in fixation duration in test trials compared with control trials (i.e., longer mean fixation) were said to have discriminated the two stimuli.

3. Results

3.1 Fixation durations

Mean fixation durations for (a) the final two habituation trials; (b) the two no-change control trials; and (c) the two test trials that presented the novel speech stimuli were calculated for each infant in each of the three conditions for approximants and vowels, and (b) and (c) are presented in Fig. 2. For each condition for each stimulus type, a 2 (age) × 3 (trial type) mixed analysis of variance was conducted, with two planned contrasts for trial type. The first contrast examined whether habituation had occurred (i.e., that there was no recovery in fixation duration for control trials compared to habituation trials). The second contrast tested whether discrimination had occurred (i.e., was there an increase in fixation duration for test trials compared to control trials)?

Results from the first contrast confirmed that for both approximants and vowels, there was no difference in fixation time between habituation and control trials (approximants, all ps > 0.2; vowels, all ps > 0.08). That is, both age groups habituated in all conditions for vowels and approximants.

More importantly, the results from the second contrast showed significant main effects for trial type, with both 6- and 9-month-old infants increasing fixation duration in test trials in the normal tilt condition for both approximants \( M_{\text{test}} = 7.2 \text{ s} \) vs \( M_{\text{con}} = 4.7 \text{ s} \), \( F(1,30) = 9.42, p < 0.01 \) and vowels \( M_{\text{test}} = 7.1 \text{ s} \) vs \( M_{\text{con}} = 4.1 \text{ s} \), \( F(1,30) = 13.81, p < 0.001 \). Thus, infants of both ages could discriminate /l/-/ɾ/ and /ɾ/-/l/ when unmodified (see Fig. 2).

![Fig. 1. (Top) Long-term average speech spectra of the approximant portion of a single token each of /ɾiː/ and /liː/. Gray line is /ɾ/, and black line is /l/. (Bottom) Long-term average speech spectra of the vowel portion of a single token each of /ɑt/ and /ʊt/. Gray line is /ɑt/, and black line is /ʊt/.](image)

![Fig. 2. Mean fixation durations for control trials (gray bars) and test trials (striped bars) for each age group in each spectral tilt condition. (Left) Approximants; (Right) vowels. Asterisks indicate significant differences.](image)
For approximants, the only other significant result was for 6-month-olds in the positive tilt condition, in which the spectral tilt emphasized the critical mid-frequency formant information between 2000 and 3000 Hz that distinguishes the two sounds. Significant main effects for trial type [F(1,30) = 9.81, p < 0.01] and age [F(1,30) = 14.94, p < 0.001], and a significant age × trial type interaction [F(1,30) = 4.66, p < 0.04] indicated that irrespective of trial type, 6-month-olds (M₀ = 7.7 s) fixated longer than 9-month-olds (M₀ = 3.9 s), and infants looked longer during test trials (Mtest = 7.6 s) compared to control trials (Mcontrol = 5.1 s). But most importantly, the significant trial type × age interaction showed that 6-month-olds had a larger increase in fixation durations during test trials than 9-month-olds. A simple effects test was significant for 6-month-olds [F(1,30) = 7.92, p < 0.01] but not 9-month-olds [F(1,30) = 2.04, p > 0.16]. Thus, younger infants successfully discriminated /l/-/l/ when high-frequency emphasis was added, whereas their older counterparts could not.

For vowels the only other significant result was for 6-month-olds in the negative condition: Significant main effects for trial type [F(1,30) = 5.97, p < 0.03] and age [F(1,30) = 4.20, p < 0.05] indicated that overall, 6-month-olds (M₀ = 5.9 s) looked longer than 9-month-olds (M₀ = 4.6 s), and fixation durations were greater in test trials (Mtest = 6.4 s) than in control trials (Mcontrol = 4.5 s). There was no age × trial type interaction, but simple effects tests showed that 6-month-olds successfully discriminated the vowels in the negative tilt condition [F(1,30) = 4.97, p < 0.04], whereas 9-month-olds did not, p > 0.2. That is, younger, but not older, infants could discriminate /l/-/l/ when emphasis was applied to the low-frequency portion of the spectrum, the location of the critical first two formants which differentiate the vowels.

3.2 Discrimination indices

Infant fixation times vary widely and because the analyses of fixation duration data shown above are based on group means, the underlying individual differences are somewhat obscured. To overcome this and examine more closely the discrimination performance of individual infants, a discrimination index (DI) was calculated for each infant in each condition:

\[
DI = \frac{\text{fixation duration}_{\text{test}}}{\text{fixation duration}_{\text{test}} + \text{fixation duration}_{\text{control}}}
\]

The greater the relative difference between an infant’s fixation durations in control and test trials, the closer the DI will be to 1. We also calculated a “discrimination criterion” (see Arterberry and Bornstein, 2002), which takes into account the variability in each age group/condition. By looking at the number of infants who scored above the discrimination criterion, we can determine the proportion of infants that performed significantly above chance: All infants whose DI was above the discrimination criterion were regarded as discriminators—those with a DI below were non-discriminators.

Figure 3 shows the DIs for all infants in all conditions. As expected, there was wide variability in discrimination performance, and in all conditions some infants in both age groups exceeded the DI criterion. Overall, the results aligned with the fixation duration results. For approximants, at least half of both age groups successfully discriminated the contrast in the normal condition (50% of 6-month-olds; 56% of 9-month-olds), but only the 6-month-olds were successful in the positive tilt condition, with 69% classed as discriminators. For vowels, 63% of 6-month-olds and 69% of 9-month-olds were discriminators in the normal speech condition, but only a majority of 6-month-olds were successful in the negative tilt condition with 69% scoring above the criterion.

4. Discussion

The results showed that both 6- and 9-month-old infants can discriminate the native contrasts /l/-/l/ and /l/-/l/ when unmodified. Six-month-old infants were also sensitive to the differences between /l/-/l/ when positive spectral tilt was applied, and to the differences between /l/-/l/ when negative tilt was applied. Nine-month-olds, on the other hand, performed best in the normal speech condition and showed little ability to discriminate the vowels or approximants when the spectral tilt was modified.

The effect of the spectral tilting was evident in the performance of the 6-month-olds. For them discrimination was facilitated by the application of an appropriate tilt: Positive for approximants, negative for vowels, because these tilts emphasized the frequency region in which the differences in the speech sounds occurred, thus making discrimination easier. However, as for fricatives, 9-month-olds were impervious to the benefits of tilting and discriminated only in the normal speech condition, revealing their native-language attunement and specificity for native speech, which occurs at this age (Kuhl et al., 2008).
Interestingly the acoustic mode of processing displayed by 6-month-olds in the fricative study who were able to discriminate /f/-/s/ in all tilt conditions was not entirely replicated. Here, the 6-month-olds could not discriminate when crucial acoustic features were obscured by spectral tilting. This may be because vowel development precedes consonant development, with infants attuning to native vowels at around 6 months of age (Kuhl et al., 1992; Polka and Werker, 1994) before attuning to native consonants at around 9 months. According to Cutler and Mehler (1993), this developmental trajectory has its basis in the infant’s “periodicity bias,” whereby young infants show a predisposition to attend to periodic sounds, such as vowels, rather than non-periodic consonant-type sounds. In this study, it seems that by 6 months, the infants were already showing evidence of attuning to the spectral profile of native periodic vowels and vowel-like approximants. Although they discriminated the contrasts when the spectral modification emphasized relevant acoustic cues, they were unable to discriminate when the less helpful spectral modification was applied.

The tendency for young infants to respond differently to vowels and consonants that have been spectrally modified is also seen in recent studies which examined infants’ discrimination of vocoded consonants and vowels. Six-month-olds could discriminate spectrally degraded consonants even when the stimuli contained only 4 channels (Cabrera et al., 2013; Cabrera et al., 2015). However, for vowel contrasts, infants only discriminated unmodified and 32-channel speech, not 16-channel speech, providing further evidence that some attunement to spectrally intact native vowels may be occurring at this age (Warner-Czyz et al., 2014).

This study confirms that 6- and 9-month-olds respond to spectrally tilted native contrasts in different ways, reflecting the developmental trajectory for speech perception that occurs during this period. Up until now, native-language attunement has been largely shown by a failure to discriminate most non-native contrasts (Best and McRoberts, 2003; Werker and Tees, 1984), but this study, our previous study (Beach and Kitamura, 2011), and other studies of spectrally modified speech (Cabrera et al., 2013; Cabrera et al., 2105; Warner-Czyz, 2014), all demonstrate that a form of attunement also occurs for acoustically modified native speech sounds. It seems that once perceptual reorganization for native speech begins to emerge between 6 and 9 months of age, infants have difficulty discriminating native speech sounds that deviate from the native spectral profile. At 6 months, infants can discriminate spectrally modified native fricatives, but their ability to discriminate modified vowels and approximants begins to wane. By 9 months, it appears that attunement to native fricatives, approximants and vowels is in place, and sounds which deviate from the native spectral profile can no longer be discriminated.

Although these results were obtained with infants with normal hearing, they may also be relevant for those seeking to optimize hearing aid prescriptions for young
infants—the findings suggest that older infants may require amplification in which the natural spectral profile is maintained as much as possible, whereas younger infants might be more adaptable to amplified speech with spectral modifications. Further work on infants with hearing loss is required before any firm conclusions can be drawn.

In future studies, it would also be useful to examine how modified spectral tilt affects speech discrimination beyond nine months of age. Perhaps once this intense period of perceptual reorganization has passed, children will be better able to accommodate modified native spectral profiles, and once again take advantage of beneficial tilts. Because attention to various spectral cues and acoustic properties of native speech continues to develop throughout childhood (Nittrouer and Lowenstein, 2007; Lowenstein and Nittrouer, 2015), it would be informative to track how spectral tilt modifications interact with these perceptual shifts until the point at which adult-like perception is reached.

Acknowledgments

This research was funded by Australian Research Council Grant DP 0559134, awarded to C.K. Thanks to Harvey Dillon, Teresa Ching, Robin Panneton, and Scott Johnson, who assisted with the stimuli; and Anna Notley, Suzy Bicanic, and Cat Grimsgaard, who assisted with recruitment and testing. Special thanks to all the babies and their parents who participated in our study.

References and links


