THE DEVELOPMENT OF PHONEME CATEGORISATION IN CHILDREN AGED 6 TO 12

Valerie Hazan and Sarah Barrett
Department of Phonetics & Linguistics, University College London, U.K.

ABSTRACT
The aim of this study was to assess the development of labelling ability across a range of phonemic contrasts acquired at different ages: /g/-/k/, /t/-/t'/, /s/-/z/ and /s/-/f/. Six-stimulus synthetic continua were created in which acoustic cues signalling these contrasts were manipulated singly or in combination. Stimuli were presented to 84 normally-hearing children aged between 6;0 and 12;6 yrs and thirteen adult controls in the form of two-alternative forced-choice identification tests using an adaptive procedure. The gradients of the identification functions increased significantly in steepness between the ages of 6 and 12 but, by 12 years old, children were still not labelling the phonemic contrasts as consistently as adults. By the age of six, there was no evidence of a greater reliance on dynamic cues such as formant transitions.

1. INTRODUCTION
It is now well established that children's ability to categorise speech sounds, as shown by their performance on identification tasks using speech continua, increases gradually over a period of years [e.g. 1,2]. This is reflected in increasingly steep identification functions and in greater proficiency with age to use the full range of acoustic cues marking a phonemic contrast.

More recently, studies by Nittrouer and others [e.g. 3] have focused on the difference in perceptual weighting given by young children and adults to cues involving rapid spectral change ('dynamic cues') relative to more static cues such as spectral shape in fricative regions. Their findings led to the Developmental Weighting Shift model [e.g. 3] which states that at an early stage of their acquisition of speech and language, children assign greater weight to dynamic cues as these signal syllable structure. They gradually shift to the weighting given by adults to acoustic cues in order to retrieve more fine-grained phonetic information as their lexicon increases. Children tested in the studies by Nittrouer and colleagues have mainly been in the age range between 3 and 8 years olds, in comparison to adults. It has been suggested that the shift away from a strong reliance on dynamic cues occurs by the age of five [e.g.4] but this evidence comes from a limited number of studies.

Although studies such as the ones quoted above have tested children in different age groups, they have typically each been limited to the evaluation of a single phonemic contrast. Therefore, it has not been possible to look at the relative development of labelling ability across a range of phonemic contrasts acquired at different ages or to evaluate the relative use of formant transition information in different contrasts in which this cue may vary in 'phonetic informativeness', to use a term promoted by Nittrouer [3]. In this study, it was therefore proposed to test children aged 6;0 to 12;6 on a set of four phonemic contrasts cued by a range of dynamic and more 'static' cues.

2. EXPERIMENT
2.1. Subjects
Listeners included 84 children (22 boys and 62 girls) aged between 6;0 and 12;6. They were divided into 6 age groups: 6;0-7;6 (15 children), 7;6-8;6 (9 children), 8;6-9;6 (7 children), 9;6-10;6 (19 children), 10;6-11;6 (16 children),11;6-12;6 (18 children). There were 13 adult controls.

Inclusion criteria were that they were: (1) native speakers of English (non-bilingual), with (2) no documented history of chronic middle-ear infection, (3) normal hearing thresholds and (4) no documented history of Specific Learning Difficulties. Most children (80/84) were pupils at six independent (private) schools in the London area. All children passed the following screening tests: (a) a test of language competence (Recalling Sentences sub-test of the CELF-R language test), (b) a pure tone audiometry screening test at 20 dB HL at frequencies between 0.25 and 4 kHz; (c) an identification test using naturally-produced versions of the minimal pairs used in the experiment.

2.2. Stimuli
The perception of four consonantal phonemic contrasts in word-initial position was tested. The minimal pairs included words taken from age-appropriate vocabulary lists that can be represented pictorially. Because of this requirement, the vocalic context could not be kept constant over all pairs. The phonemic contrasts are presented in Table 1 in expected order of acquisition given our knowledge of phonological development. Copy-synthesised versions of these words were prepared using a version of the Klatt synthesiser; each was based on an utterance produced by a phonetically-trained female speaker. A six-stimulus continuum was then created in which the acoustic patterns marking the phonemic contrast were varied in equal steps (linear for durations, logarithmic for frequency values) between the values appropriate for the two endpoints. Individual tokens were stored in individual speech files using a sampling rate of 22050 Hz.

All minimal pairs were presented in a ‘combined-cue’ condition, in which the two cues to the contrast were varying in harmony. In order to examine children's ability to integrate cues, and to evaluate their perception of the contrast when only one of the cues was present, three of these pairs were also presented in ‘single-cue’ conditions, in which only one of the two cues was varying, whilst the other was kept at a fixed value. Endpoints of all test conditions can be heard at the following web site: http://www.phon.ucl.ac.uk/home/sarahb/stimuli1.html

By comparing performance on the combined-cue condition in which both cues are present, and the single-cue condition in which only one cue provides information about the contrast, it is possible to evaluate the relative importance given to the two cues.
<table>
<thead>
<tr>
<th>Phonemic Contrast</th>
<th>Minimal pair</th>
<th>Acoustic cues varied</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɡ/-/k/ Stop voicing</td>
<td>Goat-Coat *</td>
<td>Voice Onset Time F1 transition</td>
</tr>
<tr>
<td>/θ/-/f/ Stop place</td>
<td>Date-Gate *</td>
<td>Burst frequency F2/F3 transitions</td>
</tr>
<tr>
<td>/s/-/z/ Fricative voicing</td>
<td>Sue-Zoo</td>
<td>Friction duration Voicing amplitude</td>
</tr>
<tr>
<td>/ʃ/-/ʒ/ Fricative place</td>
<td>Sue-Shoe *</td>
<td>Friction frequency F2 transition</td>
</tr>
</tbody>
</table>

Table 1. List of phonemic contrasts. The asterisked pairs were presented in both combined-cue and single-cue conditions.

Details of the parameters varied are given below:

**Goat-Coat test (/ɡ/-/k/ contrast).** In the 'combined-cue' condition, VOT ranged from 5 ms at the /f/ endpoint to 55 ms at the /k/ endpoint; F1 onset frequency varied from 380 Hz at the /f/ endpoint to 716 Hz at the /k/ endpoint. In the 'no F1 transition' condition, F1 onset frequency was kept constant at 716 Hz whilst VOT varied as in the combined-cue condition.

**Date-Gate test (/d/-/g/ contrast).** In the 'combined-cue' condition, the cues that were varied were the spectral characteristics of the initial burst transient and the F2/F3 transitions into the following vowel. The F2 onset frequency varied between 1800 Hz in /θ/ and 2522 Hz in /f/. F3 onset frequency varied between 2910 Hz in /θ/ and 2925 Hz in /f/. The burst transient was synthesised using the parallel configuration of the synthesiser and differed in the frequency and amplitude of two poles. The main change was in the frequency of F5 which varied between 5950 Hz in /θ/ and 3700 Hz in /f/ with a concomitant reduction in amplitude. In the 'Burst-cue' condition, F2 onset frequency was fixed at 2060 Hz and F3 frequency at 2916 Hz and the burst varied as above. In the 'Transition-cue' condition, the burst was fixed using intermediate values of frequency and amplitude and the F2/F3 transitions varied as in the 'Combined-cue' test.

**Sue-Shoe test (/s/-/ʃ/ contrast).** In the /s/-/ʃ/ contrast, the friction frequency and F2 transition were varied 2. The friction portion was closely modelled on the natural fricative and was made up of noise excitation passing through five forms. Three of these forms were fixed in frequency (F3: 3084 Hz; F4: 4000 Hz, F6: 6800Hz) but F4 and F6 varied in amplitude between the /s/ and the /ʃ/ endpoints (a4f: 45 to 53; a6f: 58 to 43 dB). F2 varied in frequency (1900 to 2108 Hz) and amplitude (30 to 35 dB) across the continuum and F5 varied in frequency (5400 to 4064 Hz). At vowel onset, F2 onset frequency varied between 1900 and 2108 Hz and reached a value of 2063 Hz after 40 ms. Three test conditions were prepared: in the 'combined-cue' condition, stimuli varied as above; in the 'Friction-cue' condition, parameters within the friction portion were varied but the F2 onset frequency was fixed at 2000 Hz across the continuum; in the 'Transition-cue' condition, the F2 transition varied as above and the parameters within the friction portion were kept fixed at values intermediate to those appropriate for the endpoints.

2.3. **Test procedure**

Children were tested individually in a quiet room in their school. In the test situation, these tokens were presented using a two-alternative forced-choice identification procedure. An adaptive procedure was used in order to tailor the duration of the test to a child's ability to identify the contrast. Typically, each test consisted of about 65 presentations. Testing, which included a presentation of the test conditions and various screening tests, was completed in two sessions of approximately 30 minutes. The speech perception tests were presented via a laptop computer with the child hearing the stimuli via headphones at a comfortable level. For each listening test, pictures representing the two possible choices were placed on the laptop computer and the children indicated which word they had heard by pressing the mouse button closest to the appropriate picture. At the end of each testing session, the first listening test was repeated to evaluate test-retest reliability. Each child was randomly presented with one of 5 possible orders of presentation of the nine listening tests. Adult controls were tested in a quiet room in the Department of Phonetics & Linguistics, UCL using the same procedure and equipment used with the children.

2.4. **Results**

The output of each test is an identification function that plots the percentage of one of the two alternative responses across the continuum. A maximum likelihood estimation (MLE) procedure was used to fit a probit function to each child's set of data per continuum. Two parameters were extracted to characterise each identification function: (1) the gradient of the fitted curve (expressed as probit units per stimulus number) and (2) the phoneme boundary, calculated as the 50% point of the fitted labelling curve. The identification function gradient may be used as an indicator of categorisation consistency. The phoneme boundary is an important measure for assessing effects of the various cue adjustments.

2.4.1. **Effect of age.** Analyses of variance for repeated measures (GML) were applied to the gradient data for the four combined-cue tests in order to evaluate the between-subject factor of age-group and the within-subject factor of phonemic contrast (see Figure 1). The effect of age group was strongly significant [F(6,90)=8.989; p<0.0001] and multiple pairwise comparisons showed that steeper gradients were obtained for adults than for all child groups, and that within the groups of children, gradients obtained for the youngest group (6;0-7;6) differed significantly from the 11;6-12;6 group. In order to get a more detailed picture of the change in function gradient with age for each phonemic contrast, an analysis of covariance was applied to the gradient data obtained for all children but excluding adults, with age as covariant. The effect of age on function gradient was not significant for the two voicing contrasts (/ɡ/-/k/ and /s/-/z/) but was significant for the two contrasts in place of articulation (p<0.001 for /d/-/g/; p<0.05 for /s/-/ʃ/).
2.4.2. Effect of phonemic contrast. Next, the adult and child data was analysed to look for evidence of significant difference in sharpness of categorisation across phonemic contrasts. The effect of contrast was also significant [F(3,270)=24.074; p<0.0001] but the contrast by group interaction did not reach significance. Multiple pairwise comparisons showed that a significantly steeper gradient was obtained for /d/-/g/ than all other contrasts, and that /g/-/k/ was labelled more sharply than the /s/-/z/ and /s/-/S/ contrasts, but that the two fricative contrasts did not differ significantly in mean gradient (see Figure 2).

2.4.3. Cue integration. The three contrasts (/γ/−/κ/, /δ/−γ/ and /σ/−/Σ/) that were presented in 'combined-cue' and 'single-cue' conditions were then examined in more detail. A first aim was to see whether children showed evidence of cue-integration, i.e. evidence that the combined-cue stimuli were categorised more sharply than the single-cue stimuli. For all three contrasts, there was clear evidence of cue integration as shown by a significant increase in function gradient in the 'combined-cue' condition relative to the 'single-cue' conditions (for /d/-/g/ and /s/-/f/: multiple paired comparisons: p<0.001, /g/-/k/ paired t-test p=0.002). As illustration, box-plots of gradients obtained for the three test conditions of the /d/-/g/ contrast are presented in Figure 3.

2.4.4. Performance on 'single-cue' conditions

Goat-Coat test. A perceptual effect of the F1 transition cue would be seen as a shift in phoneme boundary between two test conditions as less /g/ responses would then be expected in the absence of the F1 transition. Paired-sample t-tests were carried out on the boundary data for each of the age groups separately. The shift in boundary between the combined-cue and the 'no F1 transition' conditions was significant (p<0.04 or better) for each age group. There was no evidence of the extent of the phoneme boundary shift being greater in children than in adults (see Figure 4), hence no suggestion of a greater effect of the F1 transition in young children.

Date-Gate test. Analyses of variance for repeated measures were applied to the /d/-/g/ data for each group to examine the effect of test condition on function gradient. The effect of test condition on gradient for the Date-Gate test was significant for all listener groups (p<0.001 or better). Although sharper labelling was generally obtained for the transition-cue condition than for the burst-cue condition (see Figure 5), multiple pairwise comparisons showed that the effect was not significant in the 11:6-12:6 group, the 8:6-9:6 group and the 6:0-7:6 group. There was therefore a trend for the contrast to be more sharply categorised when cued by transition information alone than by burst information alone. However there was no evidence that this occurred to a greater extent in the younger age groups.
There was evidence in children as well as in adults of sharper phonemic contrasts continuing beyond age 12 at least. For the stop place contrast (Date-Gate), the formant transition information was found to be more informative than the burst frequency information for all groups, whilst for the Sue-Shoe contrast, steeper labelling was obtained when the contrast was cued by friction frequency information than by F2 transition information for all except two subject groups. The relative use of cues seemed therefore determined for all groups by their informativeness within a given contrast and vowel environment. These results provide further evidence of the fact that, even if as suggested by the Dynamic Weighting Shift model, a greater use is made by young children of formant transition information, the shift to adult-like use of cue information has occurred by the age of six. The process of maturation in the categorisation of phonemic contrasts, however, continues for many more years.

ACKNOWLEDGMENTS

We gratefully acknowledge the funding of the Medical Research Council (Grant G963143/4). We wish to thank the six schools which collaborated in this study: Blackheath Junior School, Channing School, City of London School for Boys, Notting Hill and Ealing Junior School, St Christopher’s School and Walker’s Primary School.

NOTES

1. A fifth contrast (Shoe-Chew) was tested in a combined-cue condition only but results are not reported here.
2. Here only F2 transition was varied in the Sue-Shoe continuum although Nittrouer [3] showed greater variation in the F3 transition. This was because the stimuli were closely modelled on specific natural utterances which did not differ in different vowel environments.

REFERENCES