Acquisition of language-specific word-initial unvoiced stops: VOT, intensity, and spectral shape in American English and Swedish

Eugene H. Buder* and Carol Stoel-Gammon†

*School of Audiology and Speech-Language Pathology, The University of Memphis, Memphis, TN, 38105, and †Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, 98105

Abstract: Children learning American English (AE) or Swedish (S) have begun to acquire language-specific aspects of word initial /t/ and /p/ by 30 months of age (1,2) and by 24 months of age (3). This report extends these language-specific observations in VOT, burst intensity, and burst spectral shape (first spectral moments coefficients of mean and standard deviation) to /k/ and /g/. The current results are obtained with revised methods (higher sampling rates) designed to reduce artifactual sources of variance and facilitate cross-age comparisons. The data suggest that language-specific patterns formerly thought to associate with distinctive place of articulation for coronal stops are actually system wide for the category of word-initial unvoiced stops. These results are consistent with the notion that children acquire specific allophonic forms quite early in development (4), even when those forms are not contrastive within the ambient language.

INTRODUCTION

Previous cross-linguistic research comparing young children’s acquisition of AE and S stops was designed to track the development of the ostensibly lamino-dental S and apico-alveolar AE coronal stop /t/ (1). Results indicated that the initial 15 ms portions of bursts produced by both adults and 30-month-old speakers from the respective languages were perceptually and acoustically distinct. VOT measures from the same tokens were also found to be language-specific. Subsequent work (4) revealed that acoustic measures from adults’ /p/ bursts were also different in the same language-specific patterns for all measures. This work utilized calibration signals in an effort to explore for artifactual effects in burst spectral shape measures, finding that although some differences could be accounted for by recording conditions, the dominant effects appeared to be reflective of true production differences. Other studies (2, 3) revealed that AE children at 30 and 24 months of age produced both types of unvoiced consonants with distinctively shorter VOTS, greater intensities, and with more “compact” burst spectral shapes (i.e., lower dispersions of energy across frequencies as measured by spectral standard deviations).

This report describes ongoing efforts to clarify the data and develop a deeper understanding of the acquisition of non-contrastive, fine-grained language-specific speech production patterns. These efforts include a) the inclusion of word initial /k/ and /g/, and b) a revamping of the methods for enhanced internal and external validity. While the long-term goal includes cross-age comparisons extending down to 24 months and upwards to include adults, the current report presents data only on productions by 30-monthers.

METHODS

Details on methods used for recording of speech, selection of tokens, and general handling of measurements are available in previous reports (1), as are details regarding calibration procedures designed to reveal the effects of variable recording conditions on spectral shape measures (4). Practical exigencies under which the prior work suffered, such as digital file storage constraints, had lead to the selection of a 20 kHz sampling rate, and the use of an analog filter (for removing low-frequency energy from recordings that would have hampered cross-language intensity comparisons). In order to establish internal validity for comparison of adults’ and 30-monthers’ data to younger children with yet higher frequency components in their stop burst spectra, a 40 kHz sampling rate protocol was adopted. In addition, to enhance the replicability and reliability of the filtering, a digital filter design was adopted in the CSpeech environment (200 coefficient FIR high-pass filtering with a 50 Hz cutoff frequency). To date, these revised methods have been applied to 454 tokens produced by 30-monthers: 89 AE /k/ and /g/, 101 AE /p/, 52, AE /t/, 71 S /k/, 90 S /p/, and 51 S /t/.
RESULTS

Descriptions of language differences and the p values of F statistics obtained by univariate ANOVAs with language as the factor are listed in Table 1.

<table>
<thead>
<tr>
<th>Word initial consonant</th>
<th>Burst intensity, relative to vowel</th>
<th>Burst spectral mean, adjusted</th>
<th>Burst spectral SD, adjusted</th>
<th>VOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>/k/</td>
<td>AE &gt; S, p = .071</td>
<td>AE &gt; S, p &lt; .05</td>
<td>S &gt; AE, p &lt; .001</td>
<td>AE &gt; S, p &lt; .001</td>
</tr>
<tr>
<td>/p/</td>
<td>AE &gt; S, p &lt; .05</td>
<td>S &gt; AE, p &lt; .05</td>
<td>S &gt; AE, p &lt; .001</td>
<td>AE &gt; S, p &lt; .01</td>
</tr>
<tr>
<td>/t/</td>
<td>AE &gt; S, p = .108</td>
<td>S &gt; AE, p &lt; .01</td>
<td>S &gt; AE, p &lt; .001</td>
<td>AE &gt; S, p &lt; .001</td>
</tr>
</tbody>
</table>

DISCUSSION

The results displayed in Table 1 indicate that language-specific differences reported previously for /p/ and /t/ are also found in /k/, with the exception of a higher spectral mean frequency for /k/ for AE than for S. The language-specific differences are otherwise in the same direction for these tokens as they had been found in adults. The results strengthen support for the interpretation that acoustic differences formerly associated with language-specific place of articulation in the coronal /t/ are actually part of a system-wide language-specific "manner" of producing word-initial unvoiced stop consonants. The pattern of early acquisition of such sub-phonemic details is consistent with the view that children acquire allophonic detail even though these details may not carry the linguistic load of signaling phoneme differences. As discussed elsewhere (2,4), the pattern of acoustic results might derive from articulatory consequences of the relatively large VOT differences (mean S: 63 ms, mean AE: 90 ms). In contrast to previous results, language-specific intensity measures were not found to be reliable. Further explorations may be needed to understand fully the consequences of sampling rate selection, room acoustics, digital filter design, and other methodological sources of influence on measurements of stop burst intensity and spectral shape.

ACKNOWLEDGMENTS

Support for this project was provided by the National Institute of Child Health and Human Development grant R01-HD32065, Carol Stoel-Gammon, PI.

REFERENCES