Speech Production-Perception Relationships in Children with Speech Delay

Kyoko Nagao¹, Mark Paullin¹, Vilena Livinsky¹, James B. Polikoff¹, Linda D. Vallino¹, Thierry G. Morlet¹, N. Carolyn Schanen², and H. Timothy Bunnell¹

¹Center for Pediatric Auditory and Speech Sciences, Nemours Biomedical Research, Wilmington, Delaware, USA
²Center for Applied Clinical Genomics, Nemours Biomedical Research, Wilmington, Delaware, USA

{nagao,paullin,livinsky,polikoff,vallino,morlet,bunnell}@asel.udel.edu, schanen@medsci.udel.edu

Abstract
This study examines the relationship between speech perception and speech production in children with speech delay (SD). Sixty-three children who participated in the Nemours Genetics of Speech Delay Project were categorized as either typically developing (TD) or SD. The children with SD were subgrouped by their articulation errors on the /s/ and /ʃ/ sounds tested in a perception experiment. An identification task was used to assess children’s perceptual ability to identify common error sounds using four sets of a 9-step synthetic continuum. The current study did not observe large boundary shifts in children with SD compared with TD children for the /s/-/ʃ/ contrast. However, one group of children who had articulation errors on both contrasting sounds exhibited a non-categorical perception pattern, whereas the other group exhibited functions similar to the control group. Furthermore, boundary shifts were observed in the perception of /k/-/g/ contrast in the children who made articulation errors on /ʃ/. The results suggest that children with delayed speech can be classified into various subgroups based on speech production as well as speech perception measures. A group of children might have poor perceptual abilities to identify phonemic sounds even though they do not exhibit articulation errors.

Index Terms: speech sound disorder, speech perception, speech production, categorical perception

1. Introduction

1.1. Speech Delay
Speech delay (SD) is characterized by difficulty producing and/or using the sounds of adult language in comparison to peers. Types of articulation errors observed in children with speech delay include sound deletions, substitutions, and distortions. Speech delay is one of the most common speech disorders among school-aged children and affects about 4% [1-3]. About 80% of children with SD require clinical intervention [4, 5]. Many children with SD also have language and learning problems. Social and educational development can be significantly disrupted in affected children and subsequent reading and learning difficulties occur in 50-70% of these children [4].

Children with SD are heterogeneous in nature [3, 6]. Some children grow out of the problem, while others may be diagnosed with developmental dyslexia or specific language impairment (SLI). The etiology of SD is yet to be discovered. However, genetic studies have identified the genes and risk loci for many speech disorders [7, 8]. Recent genetic studies have shown evidence that the risk loci for SLIs and speech sound disorders frequently overlap regions associated with dyslexia on chromosomes [9].

1.2. Speech Perception in Children with Speech Delay
A number of studies have shown a link between speech perception and speech production abilities. However, results vary for children with functional articulation disorders. Perceptual errors reflect production errors in some children with delayed speech development [10-12], while other affected children show no direct relationship between the production errors and their perceptual ability [11, 12].

1.3. Nemours Genetics of Speech Delay Project
The Nemours Genetics of Speech Delay (SpGen) Project is a collaboration between the Center for Pediatric Auditory and Speech Sciences and the Center for Applied Clinical Genomics at the Nemours Alfred I. duPont Hospital for Children, Wilmington, DE. The project seeks to more narrowly define phenotypes of SD by examining familiality in subgroups of children with SD based on acoustic, perceptual, and auditory measures. The battery of tests includes the Peabody Picture Vocabulary Test 3 (PPVT-3), Kaufman Brief Intelligence Test 2 (KBIT-2), pure-tone hearing screening, transient evoked otoacoustic emissions and their suppression, Goldman-Fristoe Test of Articulation 2 (GFTA-2) [13], a Nonword Repetition task [14], nasometry, a speech identification task, an error monitoring task, three-generation pedigree data including family history of speech and language disorders, and extraction of DNA samples obtained from the children and parents.

1.4. Preliminary Study
In [15], we reported preliminary results of perception experiments from 26 children who participated in the SpGen project. Fifteen were classified as SD and eleven were classified...
as TD based on the GFTA-2 scores. In an identification task using a 9-step synthetic continuum to assess the perception of four phonemic contrasts, the slopes of the identification functions were less steep in the SD group than in the TD group. However, unlike previous research [11], the results did not show large differences between the SD group and the TD group in category boundaries. This may be due to the heterogeneity of articulation errors among the children identified as SD in our preliminary study. That is, the articulation errors of the children in the SD group were not specifically matched with the target sounds tested in the study. In the /s/-/ʃ/ analysis, for instance, some children in the SD group may not have exhibited articulation errors on voiceless fricative sounds but only on /t/ or other sounds.

1.5. Purpose
The current report extends [15] in two ways. First, we describe categorical perception results based on 63 instead of 26 children. Second, children with speech delay are subgrouped by their articulation errors on the target sounds used in the perception experiment, and the perception results were reanalyzed by these subgroups.

2. Method

2.1. Participants
Sixty-three out of 65 children (ages 5 to 10 years) who participated in the SpGen project completed both audio recordings and perception experiments at the Speech Research Laboratory at the Nemours Alfred I. duPont Hospital for Children, Wilmington, DE. They were siblings from 33 families. At least one child in each family had been diagnosed with delayed speech development at the time of their study enrollment, but some of the identified children (proband) did not exhibit speech delay at the time of audio recording. A few of the siblings identified as typically developing during enrollment did exhibit speech delay at the time of their recording.

2.2. Stimuli
Stimuli for the categorical perception experiment are described in our previous report [15]. They consisted of four sets of 9-step continua in a CVC word frame designed to study speech perception of four contrasts: the place distinction between /s/ and /ʃ/ (suck versus shack); the /t/-/l/ manner distinction (rip versus whip); the /k/-/ɡ/ voicing distinction (cage versus gauge); and the /b/-/w/ manner (transition rate) distinction (bake versus wake). The eight words were chosen based on familiarity among young children and similar frequencies of occurrence. In addition to the experiment stimuli, 27 additional stimuli were created for a practice session.

Several repetitions of each of these words were recorded by a female native speaker of American English at the Speech Research Laboratory. These recordings were parameterized using pitch-synchronous LPC analyses and converted to residual plus line-spectral frequency parameters. Four contrasting pairs of words were selected based on similarity in their overall duration and intonation. The interpolation program then linearly compressed or expanded segment durations as well as spectral parameters to form the seven intermediate steps of each continuum. Continua were synthesized from all interpolated tokens using residual excited LPC synthesis. The four sets of continua were selected as the final stimuli based on perceived naturalness by experienced lab staff and screened by volunteer young adult listeners in identification experiments to ensure that the perceived phoneme boundary location was approximately in the middle of each continuum.

2.3. Procedure
The perception experiment was conducted in a sound-dampened room using the experiment control program Alvin [17]. Each child was seated in front of a Planar PT1701MU touch screen LCD display located next to the experimenter.

For the categorical perception test, the screen showed two pictures corresponding to the stimulus words. Children were told to listen to each stimulus carefully and select the picture corresponding to the word they heard. Stimuli were blocked by each continuum, and each stimulus in each block was randomly presented eight times. Block ordering was random across children.

Children were allowed to either use the mouse or touch the screen to select the picture. All stimuli were presented through Sennheiser HMD410 headphones. To familiarize children with the task, a practice session was conducted with the stimuli that were not used in the experiment. Children took one or more short breaks if necessary.

2.4. Subgroups based on articulation errors
Each child’s articulation was tested using the Sounds-in-Words section of the GFTA-2 [13]. Ten children were selected as the control group in this study based on their GFTA scores and age. The three oldest children who made no articulation errors were excluded in this study to reduce the effect of age. The remaining fifty children exhibited articulation errors on multiple target sounds. Table 1 lists each target sound used in the identification experiment along with the number of children who misarticulated that sound on the GFTA.

Table 1 Target sounds in the identification experiment and the number of children who misarticulated that target sound on the GFTA-2 Sounds-in-Words test.

<table>
<thead>
<tr>
<th>Misarticulated sound</th>
<th>/s/</th>
<th>/ʃ/</th>
<th>/t/</th>
<th>/l/</th>
<th>/b/</th>
<th>/k/</th>
<th>/ɡ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>29</td>
<td>18</td>
<td>30</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Since there were a good number of children who made errors on /s/ and/or /ʃ/, and these sounds were used as a contrast in the perception experiment, children were subgrouped by their articulation errors on /s/ or /ʃ/. Thirteen children who made errors on both /s/ and /ʃ/ were classified as the both-error group. Five children who had articulation errors on /ʃ/ but not on /s/ as the /ʃ/-error group, and sixteen children who had /s/ articulation errors but no /ʃ/ errors as the /s/-error group. These three subgroups were compared with the control group in the further analysis.

Mean age and range in each subgroup were 6:2 (5:0-8:2) in the both-error group, 6:5 (5:1-7:5) in the /ʃ/-error group, 7:10 (5:0-10:0) in the /s/-error group, and 7:10 (6:0-9:2) in the controls.
2.5. Analysis

To compare each subgroup with the control group at each step of the continuum, a 2-way repeated measures ANOVA (2 levels of group and 9 levels of step) was performed on the arcsine transformed percent identification scores. The current paper only reports the perception results of the two contrasts (/s/-/ʃ/ and /k/-/ɡ/) due to space limitation.

3. Results

3.1. Perception of /s/-/ʃ/

The group identification functions of the /s/-/ʃ/ continuum are shown in Figure 1. The both-error group showed a nearly linear identification function (Panel A). This suggests that the children who misarticulated both target phonemes were unable to correctly identify these stimulus sounds reliably. On the other hand, the identification functions were extremely similar between the /s/-error group and controls (Panel C). Finally, compared to the control group, the /ʃ/-error group showed a less steep slope and more variability even at the end-point stimuli (Panel B). The boundaries were similar across subgroups.

A two-way repeated measures ANOVA on the percent identification data in the both-error and control groups revealed a significant interaction effect, $F(1, 8) = 8.31, p < 0.001$, and a significant main effect for the step, $F(1, 8) = 34.44, p < 0.001$, but the main effect of group was not significant, $F(1, 8) = 1.52, p = 0.232$. A pairwise comparison revealed that the both-error group showed significantly poorer performance than the control group at all steps except step 5 (Figure 1A). On the contrary, a two-way repeated measures ANOVA on the percent identification data from the /s/-error and control groups indicated that the main effect of step was significant, $F(1, 8) = 103.30, p < 0.001$, but the interaction effect and the main effect of group were not significant, $F(1,8) = 2.46, 0.01, p_s = 0.0986, 0.743$ and, respectively. This means that the identification function in the /s/-error group was similar to the one in the control group (Figure 1C). The same analysis for the /ʃ/-error group was similar to the both-error group analysis. A significant difference was found for the main effect of step $F(1,8)=41.75, p<0.001$, and for the interaction, $F(1,8) = 2.42, p = 0.019$, but the main effect of group was not significant, $F(1,8)=1.07, p=0.320$. A pairwise comparison for each step revealed the significant difference between the /ʃ/-error and the control groups was only found for the end point stimulus (step 9), $p = 0.002$.

![Figure 1 Percentage of responses “sack” by group on the identification task with the /s/-/ʃ/continuum.](image1)

![Figure 2 Percentage of responses “cage” by group on the identification task with the /ʃ/-/ɡ/continuum.](image2)
The group identification functions of the /k/-/g/ continuum are shown in Figure 2. The both-error group showed poor performance on identifying /k/ and /g/ sounds even though the children did not misarticulate these sounds. The /s/-error group indicated poor identification of the /g/ sound, but not for /k/. These two groups exhibited a slight shift in their perceptual boundaries, while the /s/-error group did not show such a shift.

A two-way repeated measures ANOVA on the percent identification data revealed that the interactions were significant for the both-error group analysis, $F(1, 8) = 5.35, p = 0.002$, and for the /s/-error group analysis, $F(1, 8) = 2.83, p = 0.007$, but not for the /s/-error group analysis, $F(1, 8) = 1.24, p = 0.301$. For all three analyses, the main effects of step were significant, $F(1, 8) = 64.96 \sim 106.38, ps < 0.001$. The main effects of group were not significant for the three analyses, $F(1, 8) = 0.03\sim4.37, ps = 0.057\sim0.837$. A pairwise comparison revealed that the both-error group showed significantly poorer performance than the control group at stimulus steps 2, 3, 6, 7, and 9 (Figure 2A). The /s/-error group exhibited significantly different identification scores from the controls at the stimulus step 2, 6, and 9. These results suggest that the children with /s/-/g/ articulation errors exhibited poor perception to identify the end point stimuli representing /k/ and /g/ sounds (such as step 2 and 9), as well as the mid point stimuli representing highly ambiguous sounds (such as step 6).

4. Discussion

Similar to the preliminary report [15], the current study did not observe large boundary shifts in children with SD for the /s/-/g/ contrast. Contrary to our results, previous research similar to our studies reported boundary shifts [11]. This discrepancy may be explained by differences in stimulus quality. The identification test stimuli in the present study are far more natural than the highly stylistic ones used in the past. Unexpectedly, we found large boundary shifts for the /k/-/g/ contrast in subgroups of children who made /s/-/g/ articulation errors.

It is worth noting that the children who had articulation errors on both /s/ and /g/ phonemes exhibited an extremely shallow slope in the /s/-/g/ identification function. Their identification function was almost linear, which suggests that this subgroup of children did not perceive these sounds categorically. In fact, an overall low GFTA score ($M = 64.0, SD = 13.04$) suggests that the children with both /s/ and /g/ errors exhibited misarticulations on multiple phonemes. The same group of children also exhibited poor sensitivity to the /k/-/g/ contrast. These results suggest that a subgroup of children with SD had poor speech perception at both spectral and temporal domains.

Children with /s/ articulation errors, on the other hand, exhibited virtually the same identification function as the controls. An overall high GFTA score in this group ($M = 82.3, SD = 11.54$) implies that the /s/-error group did not have a severe intelligibility issue. Ongoing analysis of the misarticulated /s/ sounds suggests that this common articulation error of /s/ is a result of distortion and not substitution to a different phoneme. Hence the misarticulation does not extend to different phoneme categories in this group. Acoustic models parallel to the perceptual judgments on misarticulated sounds will be required to develop robust classification of children with speech delay.

5. Acknowledgements

This work was supported by NIDCD grant number R21-DCO007466 and Nemours Biomedical Research. The authors would like to thank Shirley Morrison for her assistance in recruiting the subjects; Jennette D. Driscoll, James T. Mantell, and Allegra Cornaglia for data collection and an early stage of data analysis; and Dr. Jobayer Hossain for his advice on statistical analysis.

6. References