Proceedings of the Institute of Acoustics

SOME ACOUSTIC CHARACTERISTICS OF SPEECH ELICITED VIA A PICTURE-NAMING TASK: A STUDY OF 6-, 8- AND 10-YEAR OLD CHILDREN

S. P. Whiteside & C. Hodgson

Department of Human Communication Sciences, University of Sheffield, UK.

1. INTRODUCTION

During the process of developing fine speech motor skills, the child is faced with the task of mastering both the timing and co-ordination of articulators in a multi-layered system, to produce gestures that are smooth and overlapping. These gestures also need to be distinct so that they can be perceptually recovered from the speech signal. The process of developing these fine motor speech skills co-occurs with the maturation of the vocal apparatus. The acoustic characteristics of adults' speech display robust sex differences with women displaying higher formant frequencies and higher fundamental frequencies [1]. It is therefore, expected that during the processes of fine motor speech development, the speech characteristics of children will also display acoustic patterns graduating towards the sex differences that are observed in adult speech [2].

Studies that have examined the acoustic phonetic characteristics of children's speech have included investigations into the coarticulation patterns of children's speech [3, 4, 5, 6, 7]. These studies have also compared the coarticulation patterns in children's speech with those of adults, and have reported inconsistent results. Some, studies have shown that coarticulatory levels of children exceed those of adult values [4, 5], while others have shown that they are lower than adult values in some phonetic contexts [3]. The inconsistent pattern of results is further highlighted by evidence that suggests that the levels of coarticulation in children's speech are similar to those of adults [6, 7]. However, it must be stated that while it has been reported that the coarticulation patterns are similar for both children (aged 3 to 7 years) and adults [6], it has also been reported that these coarticulatory effects were found to be less consistent for the children that were studied [6]. This pattern of inconsistency could be interpreted as evidence for the refinement of speech motor skills in children, and therefore, the graduation towards a skilled adult motor speech system.

Numerous studies investigating the acoustic phonetic developmental characteristics of children's speech have focussed on the acoustic changes that occur with increasing age [2, 8, 9, 10, 11]. The acoustic changes that have been reported have included changes in the patterns of fundamental frequency and formant frequencies [2, 8, 9, 10, 11]. The general trend in these studies has been for fundamental frequency and formant frequencies to drop with increasing age, and for female children to have higher fundamental and formant frequencies than their male peers. In addition there is evidence that male children exhibit more marked drops in frequency than their female peers [10].

This study investigates some acoustic phonetic characteristics of the speech of 20 pre-adolescent (6-, 8- and 10-year olds) boys and girls. The speech data were elicited via a picture-naming task. This investigation examines the acoustic and phonetic characteristics of vowels and consonants of some of this speech data. Acoustic phonetic characteristics included a selection of formant frequency values, coarticulation (or gestural overlap) patterns and temporal patterns. These temporal patterns consisted of total utterance duration, total speaking time and total pause duration. The acoustic phonetic data were examined to determine the nature and extent of developmental patterns, and also the nature and extent of sex-related patterns. The results are discussed with reference to the developmental patterns of speech and the sexual dimorphism of the vocal apparatus [2].

2. METHOD

2.1 Subjects

Three boys and three girls from two age bands: 6 years (mean age 6.03 yrs) and 8 years (mean age 8.18 yrs), and four boys and four girls aged 10 years (mean age 9.97 yrs), served as subjects in the study. The 20 children were all from the same local primary school in the north east of England (Tyneside). All children had lived in the local area all their life and spoke English as their first language. None of the 20 children had any speech, language or hearing difficulties.

2.2 Data collection

The recordings were made in a quite room using a portable AIWA HDS1 DAT recorder and a stereo microphone. Picture cards were used to elicit a possible total of nine target phrases from the children, (*The red/blue/green bar/jar/car*). Also included in the elicitation, were four distracters, for example, *The red boat*, *The green balloon*. Prior to the actual collection of the data, children were asked to name the colours and the pictures separately. This was done to ensure that none of the children were colour blind and also to establish picture names and therefore avoid confusion during the recording session

2.3 Acoustic Analyses

The elicited data were digitised with a sampling rate of 10 kHz using a Kay Computerised Speech Lab (CSL- Model 4300). All acoustic analyses were performed on the KAY CSL using routines, which are outlined below.

2.3.1 Schwa formant frequencies. F2 values for schwa were obtained for schwa in *The*, for the phrases beginning *The red*.... and *The blue*....., using LPC analysis. A frame length of 10ms was used and the selected point was the midpoint of schwa, which was determined by the speech pressure waveform. Measurements were validated using wideband FFT spectrograms.

2.3.2 Formant frequencies (F2 and F3) of /r/. The midpoint of /r/ was located in *red*, with the help of time-aligned wide-band spectrograms. F2 and F3 values were obtained using the same method outlined above for schwa (2.3.1.).

It has been shown that in the case of retroflex consonants, F3 is a more reliable indicator of coarticulation than F2 [12, 13]. Therefore in order to examine the degree of coarticulation of /r/ with /ɛ/, F3 values, in addition to those of F2, were obtained for /r/.

2.3.3 Formant frequencies (F2 and F3) of /ɛ/. The midpoint of /ɛ/ and was located in *red* using wide-band spectrograms. F2 and F3 values were obtained using the same method outlined above for schwa.

2.3.4 Formant frequencies (F2) of the onset and midpoints of /ɑː/. Formant frequency values for the onsets and mid points of the vowel /ɑː/ in the context *bar* were obtained as follows. F2 values were taken at the onset of the vowel and F2 values were taken at the temporal midpoint of the vowel (half way through the vowel from the onset of the vowel).

2.3.5 Phrase and 'pause' durations. The durations were measured for each phrase by locating the start and the end of the speech pressure waveform. In addition, the duration of any acoustic closures (i.e. no acoustic activity) after 'The' and before *bar/jar/car* were measured, and labelled as 'pauses'. The 'pause' durations of both positions were added together to give the total 'pause' duration. This total 'pause' duration was then subtracted from the phrase durations to give a total speaking time. All the aforementioned measurements were made in milliseconds.

2.3.6 Coarticulation Patterns. Measures that were taken to investigate some coarticulation patterns included:

- Differences between the F2 values of /ɛ/ and the F2 values of /r/ (Hz), where patterns of greater F2 differences could be interpreted as evidence for less coarticulation.
- Differences between the F3 values of /ɛ/ and the F3 values of /r/ (Hz), where patterns of greater F3 differences could be interpreted as evidence for less coarticulation, and differences between the F2 onset and midpoint values of /ɑː/ (Hz) in the context *bar*.

3. RESULTS

3.1 Acoustic phonetic parameters

3.1.1 Schwa formant frequencies. The mean F2 values for schwa are given in Table 1 by sex and age for two contexts (schwa preceding: /r/ in *red*, /b/ in *blue*). The F2 data for both contexts were subject to a two way (by age and sex) ANOVA. The data for schwa preceding /r/ indicated that there were sex differences ($F(1, 48)=45.8$, $p<.0001$), with the male children having lower formant frequencies than their female peers (see Figure 1). There were also significant age differences ($F(2, 48)=8.5$, $p<.005$) with a general drop in frequency from the age of 6 years to 10 years (see Figure 1). No age-by-sex interactions, however, were found. A post-hoc Scheffé test indicated that the age differences were significant ($p<.05$) between the data of the 10- and 8-year olds, and those of the 10- and 6-year olds. No significant differences were found between the data of the 6- and 8-year olds.

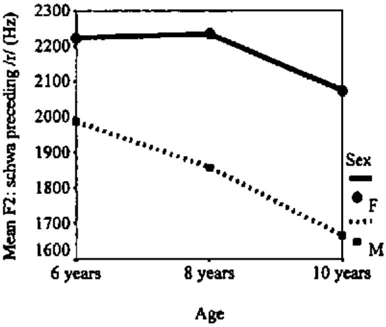

Figure 1: Mean F2 values for schwa preceding /r/ (Hz) by age and sex

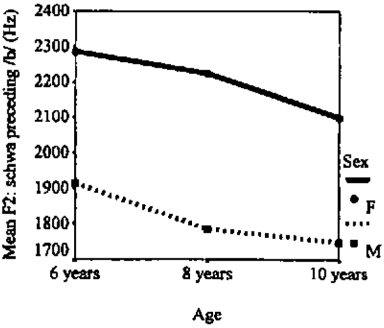

Figure 2: Mean F2 values for schwa preceding /b/ (Hz) by age and sex

The F2 data for schwa preceding /b/ indicated significant age differences ($F(2, 48)=3.3$, $p<.05$), with a general drop in frequency with age (see Figure 2). Significant sex differences ($F(1, 48)=44.9$, $p<.0001$) were also found with the male children displaying lower formant frequencies than their female peers. Pairwise comparisons of estimated means indicated significant age differences ($p<.05$) between the data of the 6-year olds and those of the 10-year olds, with. No significant differences were found between the data of the 6- and 8- year olds. No significant age-by-sex interactions were found.

Parameter	Age 6 yrs.		Age 8 yrs.		Age 10 yrs.	
	F	M	F	M	F	M
F2: schwa preceding /r/ (Hz)	2223.5	1987.2	2234.2	1846.4	2073.9	1665.1
F2 of schwa preceding /b/ (Hz)	2286.5	1913.3	2224.2	1785.7	2099.3	1748.0

Table 1: Mean F2 values (Hz) for schwa according to both sex and age.

3.1.2 Formant frequencies (F2 and F3) of /r/. The mean F2 and F3 values for the mid point of /r/ are given in Table 2 according to both sex and age group. A two-way ANOVA (by age and sex) for the F2 data indicated that were no significant age differences. There were however, significant sex differences ($F(1, 54)=15.5$, $p<.0001$), with the female children exhibiting higher F2 values than their male peers. No age-by-sex interactions were found for the F2 data.

When the F3 data were subjected to a two-way ANOVA (by age and sex), similar trends to the F2 data were observed. Again, no significant age differences or age-by-sex interactions were found. However, significant sex differences were found ($F(1, 54)=4.9$, $p<.05$), with the female children's data generally exhibiting higher F3 values than that of their male peers.

Parameter	Age 6 yrs.		Age 8 yrs.		Age 10 yrs.	
	F	M	F	M	F	M
F2 of /r/ in *red* (Hz)	1695.6	1577.8	1727.8	1496.8	1738.8	1404.1
F3 of /r/ in *red* (Hz)	2581.4	2616.1	2562.9	2317.3	2709.8	2323.1

Table 2: Mean F2 and F3 values (Hz) for /r/ by sex and age.

3.1.3 Formant frequencies (F2 and F3) of /ɛ/ and F2 onset and F2 midpoints of /ɑː/ in *bar*. The mean F2 and F3 values for /ɛ/ are given in Table 3 according to both sex and age group. A two-way ANOVA (by age and sex) for the F2 data indicated that were significant age differences ($F(2, 52)=7.3$, $p<.005$). Further to this, a post-hoc Scheffé test indicated that there were significant differences ($p<.05$) between the F2 data of the six year olds and those of the 10 year olds, with the six year olds having higher F2 values. There were also significant sex differences ($F(1, 54)=15.5$, $p<.0001$), with the female children exhibiting higher F2 values than their male peers. No age-by-sex interactions were found.

When the F3 data were subjected to a two-way ANOVA (by age and sex), results indicated that were significant age differences ($F(2, 52)=10.1$, $p<.0001$). Further to this, a post-hoc Scheffé test indicated that consistent with the F2 data, there were significant differences ($p<.05$) between the F3 data of the six year olds and those of the 10 year olds, with the six year olds having higher F3 values. No significant sex differences were found for the F3 data, however, there was a significant age-by-sex interaction ($F(2, 52)=6.7$, $p<.005$). This interaction took the following form: both the male and female children exhibited a marked fall in the frequency of F3, between the ages of 6 and 8 years. However, between the ages of 8 and 10 years, the girls exhibited a slight rise in the frequency of F3, whereas in contrast to this pattern, the boys exhibited a marked fall in the frequency of F3.

When the F2 onset values for *bar* (see Table 3) were subjected to a two-way ANOVA (by age and sex), results indicated that were significant age differences ($F(2, 49)=9.2$, $p<.0001$). Further to this, a post-hoc Scheffé test indicated that there were significant differences ($p<.05$) between the F2 onset data of the six year olds and those of the 10 year olds, and also between the data of the 10 year olds and those of the eight year olds. In both comparisons, the 10-year olds had the lowest F2 values. In addition, significant sex differences were found ($F(1, 49)=47.0$, $p<.0001$) however, there was no significant age-by-sex interaction. The F2 mid point values for *bar* (see Table 3), were also subjected to a two-way ANOVA (by age and sex). Results here indicated that were both significant sex ($F(1, 49)=39.2$, $p<.0001$) and age differences ($F(2, 49)=8.5$, $p<.005$). Further to this, a post-hoc Scheffé test indicated that the between group significant differences ($p<.05$) replicated those found for the F2 onset *bar* data as outlined above. In both comparisons, the 10-year olds had the lowest F2 values. No significant age-by-sex interaction was found for these data.

Parameter	Age 6 yrs.		Age 8 yrs.		Age 10 yrs.	
	F	M	F	M	F	M
F2 of /ɛ/ in *red* (Hz)	2389.3	2371.2	2395.7	2247.8	2328.8	2087.3
F3 of /ɛ/ in *red* (Hz)	3582.6	3714.1	3305.2	3489.4	3404.8	2879.3
F2 onset /ɑː/ (*bar*)	1642.0	1474.0	1638.9	1384.3	1508.4	1287.6
F2 mid /ɑː/ (*bar*)	1534.9	1381.6	1538.1	1381.4	1426.1	1269.3

Table 3: Mean F2 and F3 values (Hz) for /ɛ/ and F2 onset and mid points of /ɑː/ (*bar*) by sex and age.

3.1.4 Phrase and 'pause' durations. The mean values for the total phrase duration, total 'pause' duration and the total speaking time are given in Table 4 according to sex and age. The total phrase duration, total 'pause' duration and the total speaking time data were subjected to a two way (by age and sex) ANOVA. Results indicated that there were significant age differences for all three parameters of total phrase duration ($F(2, 155)=37.2$, $p<.0001$), total 'pause' duration ($F(2, 155)=20.4$, $p<.0001$) and total speaking time ($F(2, 155)=45.1$, $p<.0001$), with a general decrease in duration with increasing age. Figures 3 and 4 illustrate this trend for the total 'pause' duration and total speaking time, respectively. Post hoc Scheffé tests indicated that there were significant age differences between the data of the 6-year olds and the 8-year olds, and between those of the 6-year olds and the 8-year olds. No significant differences were found between the data of the 8- and 10-year olds. Significant sex differences were found for total speaking time ($F(1, 155)=4.3$, $p<.05$), with the female children displaying longer speaking times (see Figure 4). No sex differences were found for either total 'pause' duration, or total speaking time, however. In addition, no significant age-by-sex interactions were found for any of the three parameters. When the three temporal parameters were subjected to a Pearson's product moment correlation coefficient test, significant correlations were found for the following pairs: total phrase duration and total 'pause' duration ($r= .94$, $p<.0001$); total phrase duration and total speaking time ($r= .81$, $p<.0001$) and total speaking time and total 'pause' duration ($r= .56$, $p<.0001$).

Parameter	Age 6 yrs.		Age 8 yrs.		Age 10 yrs.	
	F	M	F	M	F	M
Total phrase duration (ms)	1533.2	1516.8	1006.5	839.2	889.0	853.7
Total 'pause' duration (ms)	460.4	431.7	109.4	51.8	70.5	103.5
Total speaking time (ms)	1072.7	1085.1	897.1	787.5	818.4	750.2

Table 4: Mean values (ms) for total phrase duration, total 'pause' duration and total speaking time by age and sex.

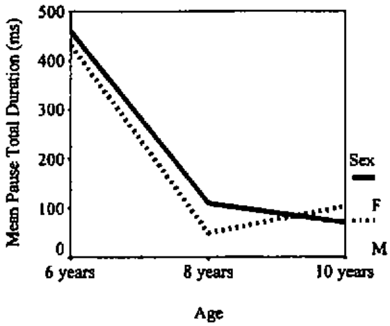

Figure 3: Mean total length of pauses (ms) by age and sex.

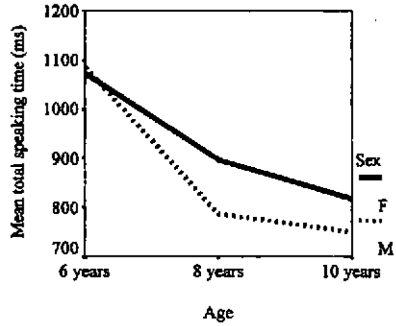

Figure 4: Mean total speaking time (ms) by age and sex.

3.2 Coarticulation Patterns

3.2.1 F2 and F3 change between mid /r/ and mid /ε/. The values for mean F2 and F3 change (Hz) between the midpoint of /r/ and the midpoint of /ε/, are given in Table 5 by sex and age. The F2 differences were subjected to a two-way (by age and sex) ANOVA, results indicated no significant differences were found for any of the comparisons. When the F3 differences were subjected to a two-way (by age and sex) ANOVA, a different picture emerged. Results showed that although no significant sex differences or age-by-sex interactions were found, there were significant age differences ($F(2, 48)=5.4$, $p<.01$). A subsequent post-hoc Scheffé test indicated that there were significant differences between the data of the 6-year olds and the 10-year olds, with the 6-year olds showing greater F3 differences.

3.2.2 F2 change between onset and midpoint of /ɑ:/ in *bar*. The values for mean F2 difference between the onset and midpoint of the vowel /ɑ:/ in *bar* are given in Table 6. The F2 differences were subjected to a two-way (by age and sex) ANOVA. No significant differences were found for any of the comparisons.

Parameter	Sex	Age 6 yrs.	Age 8 yrs.	Age 10 yrs.
F2 difference between midpoints of /r/ and /ε/	F	-595.8	-667.9	-590.0
	M	-792.6	-767.0	-683.2
F3 difference between midpoints of /r/ and /ε/	F	-1032.2	-742.3	-695.0
	M	-1100.4	-1222.4	-556.3

Table 5: Mean F2 and F3 differences between mid points of /r/ and /ε/ (Hz)

Parameter	Sex	Age 6 yrs.	Age 8 yrs.	Age 10 yrs.
F2 change: /ɑ:/ mid vowel minus /ɑ:/ onset (*bar*)	F	-107.1	-100.8	-82.3
	M	-92.2	4.1	-18.3

Table 6: Mean F2 difference (Hz) between the mid point and onset of /ɑ:/ in *bar* by age and sex.

4. DISCUSSION

The data reported here show evidence of sex differences in the formant frequencies of the 6-, 8- and 10-year olds in this study, with the female children having higher formant frequencies than their male peers. Moreover, the evidence appears to be robust, because these patterns were found for the data of the alveolar approximant /r/, the vowel /ɛ/ and both the onset and mid point values of the vowel /ɑː/ (see Table 6). These sex differences in the patterns of formant frequencies replicate those of adults [1, 14] and could be a result of the females having smaller vocal tracts than their male peers. It would therefore appear that there is already evidence of robust sex differences in the formant frequencies of pre-adolescent children, a finding that replicates those of previous developmental studies [9, 10, 11]. What is perhaps more striking in the data reported here is the age-by-sex interaction that was found for the F3 data of the vowel /ɛ/. Here, the F3 data for the male children exhibited a significant fall in frequency between the ages of 8 and 10 (see Table 2 and Figure 1), whereas those of the female children exhibited a slight rise between these ages. This could be interpreted as some evidence for marked changes in the vocal tracts of the male children between these ages.

The sample of coarticulation patterns reported here indicated that there were no significant age, sex or age-by-sex differences for the F2 differences between the mid points of /r/ and /ɛ/ gestures. There were however, significant age differences for the F3 differences between these two gestures. These findings suggests that F2 were not be sensitive enough to detect the gestural overlap of retroflexed consonants, a finding which has been reported elsewhere [12, 13]. In addition, the age differences in the F3 data suggest that the 6-year old children show less overlap than both the 8-year olds and the 10-year olds, between the gestures of alveolar approximant and the vowel. The pattern of no significant age, sex or age-by-sex differences for the F2 differences of the vowel /ɑː/ in the context *bar*, suggest that all three age groups and males and females displayed similar patterns of gestural overlap for the /bɑː/ gestures. These findings suggest that the development of coarticulation patterns is sensitive to phonetic context. It is suggested that because some gestures involve more complex and secondary articulations (such as those that are involved in the alveolar approximant /r/), are mastered later than those that do not (e.g. in the case of /b/). On this premise, it is therefore proposed that the former types of gesture will display less gestural overlap with neighbouring gestures in the speech of the 6-year olds because they have a less skilled and therefore, less flexible and accommodating articulatory system, which is still undergoing development.

The data for total utterance duration, total speaking time and total pause duration displayed significantly longer durations for the 6-year olds compared to the 8- and 10-year olds. This pattern of longer durations in total speaking time in the speech of the youngest children could be interpreted as evidence for motor speech development, where we see faster articulation with increased levels of skill. In addition, the longer inter-word pauses in the speech of the youngest children suggest word by word chunking in the planning and production of the phrases. In contrast, the 8- and 10-year olds have few pauses and are therefore possibly adopting phrase-size chunking in the planning and production of the phrases.

These preliminary data are indicative of both the age and sex differences that co-occur during speech development. The data for sex differences in formant frequencies replicate those of adult studies [1, 14] and indicate that these differences are present before puberty. The coarticulation data suggest that phonetic context is an important factor when investigating the developmental patterns of coarticulation, in that we may find a range of age-related differences in different gesture sequences. In addition, there is evidence from the temporal patterns that the 6-year olds adopted different speech encoding strategies in the picture-naming task, compared to the 8- and 10-year olds.

5. ACKNOWLEDGMENTS

We wish to thank the Head teacher, staff, and pupils of Priory Primary School, Tynemouth, Tyneside, for their help and co-operation with this study.

6. REFERENCES

1. Peterson, G., and Barney, H. "Control methods used in the study of vowels", *Journal of the Acoustical Society of America,* 24, 1952, 629-637.
2. Kent, R. D. "Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies", *Journal of Speech and Hearing Research*, 19, 1976, 421-446.
3. Goodell, E. W., and Studdert-Kennedy, M. "Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: a longitudinal study", *Journal of Speech and Hearing Research,* 36, 1993, 707-727.
4. Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. "The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children", *Journal of Speech and Hearing Research*, 32, 1989, 120-132.
5. Repp, B. "Some observations on the development of anticipatory coarticulation", *Journal of the Acoustical Society of America,* 79, 1986, 1616-1619.
6. Sereno, J. A, Baum, S. R., Marean, G. C., and Lieberman, P. "Acoustic analyses and perceptual data on anticipatory coarticulation in adults and children", *Journal of the Acoustical Society of America,* 81, 1987, 512-519.
7. Turnbaugh, K., Hoffman, P., Daniloff, R. D., and Absher, R. "Stop-vowel coarticulation in 3-year olds, 5-year olds and adults", *Journal of the Acoustical Society of America*, 77, 1985, 1256-1258.
8. Bennett, S. "A 3-year longitudinal study of school-aged children's fundamental frequencies", *Journal of Speech and Hearing Research,* 26, 1983, 137-142.
9. Busby, P. A., and Plant, G. L. "Formant frequency values of vowels produced by preadolescent boys and girls", *Journal of the Acoustical Society of America*, 97, 1995, 2603-2606.
10. Hasek, C. S., Singh, S., and Murry, T. "Acoustic attributes of preadolescent voices", *Journal of the Acoustical Society of America,* 68, 1980, 1262-1265.
11. Eguchi, S., and Hirsh, I. "Development of speech sounds in children", *Acta Otolaryngologica*, 257, 1969, 5-48 (Suppl.).
12. Krull, D., and Lindblom, B. "Coarticulation in apical consonants: acoustic and articulatory analyses of Hindi, Swedish and Tamil", Swedish Phonetics Conference, Nässlingen, 29-31 May 1996, TMH-QPSR 2/1996, 73-76.
13. Krull, D., Lindblom, B., Shia, B-E., and Fruchter, D. "Cross-linguistic aspects of coarticulation: an acoustic and electropalatographic study of dental and retroflex consonants", *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 3, 1995, 436-439.
14. Childers, D. G., and Wu, K. "Gender recognition from speech: Part II. Fine analysis", *Journal of the Acoustical Society of America,* 90, 1991, 1841-1856.