Acoustics of children's speech: Effects of age on duration, pitch and formants

Sungbok Lee*, Alexandros Potamianos and Shrikanth Narayanan

AT&T Labs–Research, 180 Park Ave, Florham Park, NJ 07932-0971, U.S.A.

June 12, 1997

To be submitted to the *Journal of the Acoustical Society of America*

Running title: *Acoustics of children's speech*

Received: __________

*The author's current address is Central Institute for the Deaf, St. Louis, MO 63110, U.S.A.

Abstract

PACS number: 43.72.Ar

INTRODUCTION

As children grow older after the acquisition of basic phonological knowledge of a language, acoustic properties of normal children's speech change toward those of adults, as the vocal organs grow and speech production skill improves by experience. Therefore, "speech development" is a combined and progressive process in which both the anatomical growth of the vocal organs, which governs the magnitude of spectral parameters such as pitch and formant frequencies, and the neuromuscular control of the speech apparatus, which determines timing and precision in speech production, are integrated. Naturally, one important issue in the study of speech development is to know how the magnitude and variability of pitch and formants and duration vary with age and when they reach adult range. The purpose of this study is to explore the magnitude and variability issues of children's speech based on the acoustic analyses of a recently collected speech database (ref).

For the purpose, mgnitude and variability of duration, pitch, and formant frequencies, together with the spectral variability of the whole spectrum, are investigate as a function of age and gender. As the database enables a cross–sectional study of children speech, with respect to adults, from ages five through eighteen with resolution of one–year of age, it is possible to get detailed trends on speech development with age in both temporal and spectral domains. Such systematic knowledge on the age–dependent patterns of the acoustic parameters obtained from the present study is also valuable for speech applications such as automatic recognition of children's speech (Wilpon and Jacobson, 1996; Potamianos et al., 1997), evaluation and training of deaf children's speech through computer–interface (ref), and text–to–speech synthesis (ref). The knowledge is also important for associating the acoustic development in children with the underlying anatomical development of the vocal organs, which is essential for the creation of a better developmental model of the vocal tract for speech production research (Goldstein, 1980).

Specifically, duration is measured on vowels, /s/ in inter–syllabic level, and meaningful sentences. Pitch and formants are measured on monophthongal vowels. Variability of the acoustic parameters is measured in terms of between– and within–subject variations. The results of magnitude and variability measurements are discussed in terms of the anatomical developments of the vocal tract and the degree of maturity of speech motor control. The relationships between the magnitude and the variability among the three acoustic parameters are also explored as a function of age and gender. This paper is organized as follows: The database analyzed in this study is briefly reviewed in Section I.. In Section II., measurement procedures used to estimate the acoustic parameters are described. The results are presented in Section III.. Discussion of the results of the current study, together with the comparison with those of previous studies, are given in Section IV. followed by a conclusion. The pitch and formant data for vowels are listed in Appendix as a function of age and gender.

I. DATABASE

The database analyzed in the present study was collected from 436 children of ages five to eighteen, with a resolution of one-year of age, and 56 adult reference speakers (Miller et al., 1996). The speech material consisted of the ten monophthongal and five diphthongal vowels of American English and of five meaningful sentences. Target words for the monophthongs analyzed in this study were *bead (IY), bit (IH), bet (EH), bat (AE), pot (AA), ball (AO), but (AH), put (UH), boot (UW), and bird (ER)*. The target words were produced in the carrier sentence “I say uh – again” except for children of age five and six who produced target words only. The five sentences were: 1) “He has a blue pen.” 2) “I am tall.” 3) “She needs strawberry jam on her toast.” 4) “Chuck seems thirsty after the race.” 5) “Did you like the zoo this spring?”. All utterances were produced twice from each speaker.

Recordings were made in a sound-treated booth located inside a glass/panel enclosure, using a high–fidelity microphone (Bruel & Kjaer model #4179) connected to a real–time waveform digitizer (Desklab–216, Gradient Technology Inc.) with 20 kHz sampling rate and 16–bit resolution. Target utterances were presented in random order. For subjects that couldn't read a given target utterance, a pre-recorded utterance was played and the subject was asked to repeat it. The recording facility was located in the St. Louis Science Center, Missouri.

After the data were collected, each waveform file was examined by listening to the recorded speech. Waveform files that were truncated or of very low recording quality were marked as “chopped” or “bad” and discarded. From the initial 24630 waveform files, 24152 files were judged to be good. A detailed description of subjects, speech material, data collection procedures, and recording environment can be found in Miller et al. (1996).

II. MEASUREMENTS

In this section, the automatic procedure used to measure the duration, pitch and formant frequencies is presented and evaluated by comparing its results to corresponding hand–measured data points for a small subset of the data. The automatic segmentation procedure used to obtain the phonemic, word and sentence boundaries is also outlined.

A. Preprocessing of database

For the analysis of the database, it was necessary to determine the boundaries of each phoneme in a given utterance. In order to process the large number of speech samples within a reasonable amount of time, the segmentation and phonetic labeling of each waveform was automatically computed by aligning the analysis speech frames to the states of the corresponding phonemic units [8]. Hidden

Markov models (HMMs) were built for each phone unit from an adult speaker population and an initial segmentation was obtained for the all sentences in the children database using probabilistic alignment. Next, using the initial segmentation, phone HMMs units were trained from the children database and segmentation was recomputed using the new acoustic models. The automatic segmentation procedure produces a label file in which the beginning (end) of each phoneme corresponds to the entry time to the first state (exit time from the last state) of the HMM phonemic unit. Time marks have a 10–msec resolution which is the rate of the analysis frame.

The accuracy of the automatic segmentation was evaluated using hand–measured durations of 160 randomly selected vowel segments from sixteen subjects of ages five, eight, eleven, and adults. The results are shown in Fig. 1(a). Mean difference between the automatically computed and the hand–measured values is −17.5 msec with standard deviation of 37.0 msec.

B. Duration

Durations of the vocalic segments of the target words for the ten monophthongs were computed from the automatic segmentation as the difference between the beginning and the end of the speech unit under consideration. The average duration of the fricative portion of /s/ in the carrier sentence "I [s]ay – " was also computed (40 occurrences per speaker). Finally, the duration of the five sentences were estimated from the word-level automatic segmentation. Since the average duration varied significantly among sentences it was deemed necessary to normalize sentence duration with respect to average duration (computed from the adult reference speakers).

As it can be seen in Fig. 1(a), vowel durations were systematically underestimated by the automatic procedure and thus vowel durations obtained in this study may be somewhat shorter than their actual values. However, the automatic segmentation of /s/ and the end–point detection of each sentence were quite accurate within the limit of resolution (i.e., 10 msec) of the automatic segmentation procedure.

C. Pitch and formant frequency

The fundamental frequency (F0) and the first three formants frequencies (F1, F2, F3) of the ten monophthongs were estimated using an automatic pitch and formant tracking program in the ESPS signal processing package by Entropic Research Laboratory Inc., which utilizes a dynamic programming technique to choose best pitch and formant tracks among candidates (Secrest and Doddington, 1983). Speech was downsampled to 12 kHz and processed using a 12 msec Hamming window (updated every 5 msec), a first–difference preemphasis factor of 0.94, and linear prediction analysis of order 12. Changes of the analysis parameters were found to have little effect on performance.

The performance of the automatic program was evaluated from hand–measured pitch and formant values of 96 randomly selected vocalic segments from sixteen subjects of ages five, eight, eleven, and adults. The manual estimation process was as follows: 1) three 20 msec segments were selected around the steady state portion of each vocalic segment, 2) the pitch and formants of each segment were measured by visual inspection of the corresponding FFT spectrum, spectrogram, and

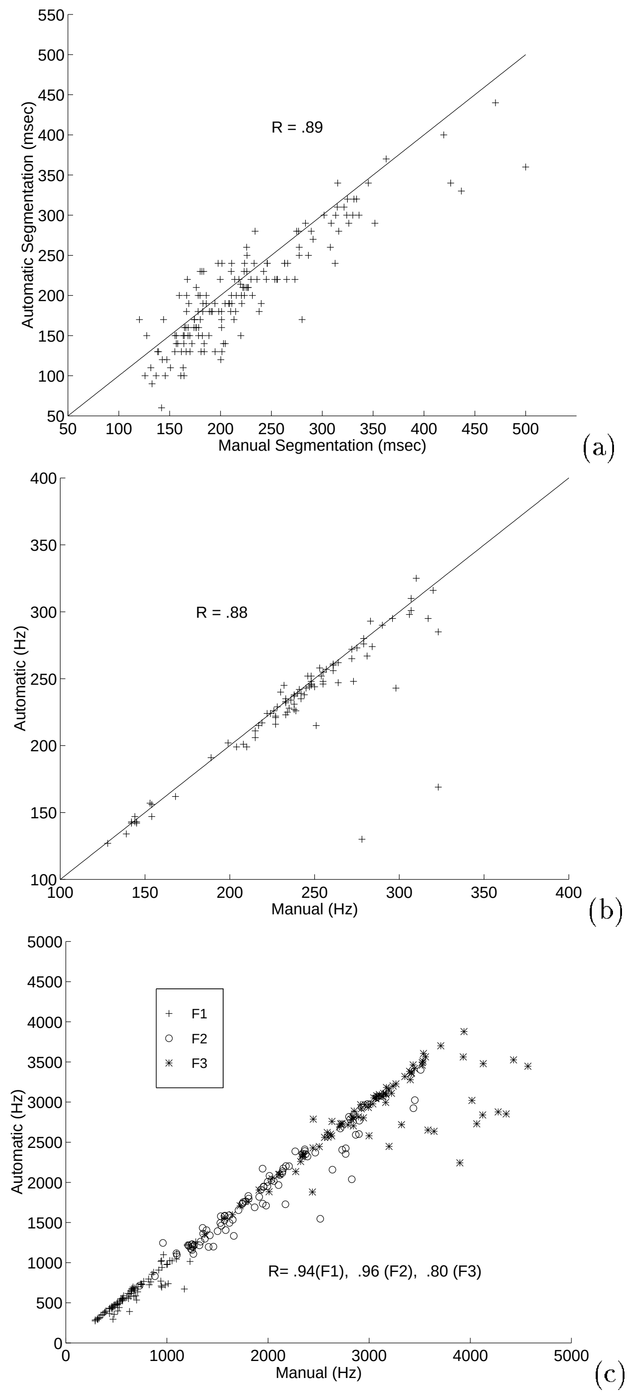


Figure 1: Comparison of manual and automatic measurements of (a) vowel duration, (b) pitch (F0), and (c) F1, F2 and F3.

the poles of the LPC spectral envelope, 3) the hand–measured values of pitch and formants were averaged over the three sub–segments. The manually estimated pitch and formant values were compared to the corresponding median values of the automatically estimated pitch and formant tracks. The results are shown in Fig. 1(b) for pitch and in Fig. 1(c) for formant estimates. Mean differences (standard deviation) between the manually and automatically estimated values in Hz were −7.6 (23.8) for F0, −43.6 (87.2) for F1, −92.4 (183.8) for F2, and −193.1 (400.7) for F3. Pitch tracking by the automatic program was fairly reliable across age and gender except occasional pitch halvings, as can be noticed from Fig. 1(b). The program also yielded relatively accurate estimation of the first formant frequency. However, it had troubles in tracking F2 and F3 for young children due to the poor spectral resolution (caused by high fundamental frequency values), spurious spectral peaks and formant tracks merging. Similar problems were encountered when formant frequencies were estimated by visual inspection of the speech spectrogram. In general, the measurement error for both the manual and automatic formant estimates increases with fundamental frequency (ref).

As some erroneous pitch and formant values are expected in the raw pitch and formant data, the raw data were refined using the following procedures: the initial formant data were grouped by vowel, age and gender (10 vowels x 15 age groups x 2 genders) and outliers in each group were removed using a two–sigma ellipse. Next, each data file was visually examined based on the mean and standard deviation of the data file and whenever one of the F1-F3 values was judged to be too low or too high, the corresponding formant set was discarded. Despite our efforts to remove erroneous formant values, some might have escaped the visual inspection process. It appears that the refined formant data set included some underestimated F2 and F3, especially for children age eight and lower. For pitch estimates, as can be noticed from Fig. 1(b), the large mean difference and standard deviation of F0 estimation error was due to a few occurrences of "pitch halving" by the automatic program. These erroneous pitch values were removed by a procedure similar to the one used for the refinement of the formant data. These refined pitch and formant data are presented in this paper.

D. Spectral envelope variability

The smooth spectral envelope or equivalently the cepstrum coefficients derived from the spectral envelope are the most common set of features used in automatic speech recognition and are extensively used in speech coding and speech synthesis (ref). The spectral envelope can be used to measure spectral variability of speech sounds in a robust way.

In this paper, two measures of spectral variability are computed: 1) mean square difference of average spectral envelopes between two repetitions of the same target vowel, and 2) mean square spectral difference between the first and second half segments within a vowel. The former is compared with the within–subject formant variability. The later is interpreted as the amount of total spectral change between the front and the back part of the vowel.

The spectral envelope of a given vowel segment was computed using a mel–frequency filterbank of 29 frequency–bands (spanning 100 to 6000 Hz). The cepstrum coefficients (twelve cepstrum coefficients plus energy) were computed from the log spectral envelope using the inverse cosine transform. Spectral distance was computed as the Euclidean distance between two sets of cepstral coefficients.

E. Variability measurements

Variability of duration, pitch and formants associated with each age group were measured in terms of within– and between–subject variations. For a given parameter, between–subject variation for an age group was computed as the standard deviation of the subject mean value for all subjects in the age group.

Since each subject produced a given target utterance twice, within–subject variation was computed as the difference of the magnitude of a given parameter between two repetitions. Group intra–subject variation was defined as the average intra-subject variation of that age group. Finally, the coefficient of variation (COV) was computed as the ratio of between– or within–subject variation over the corresponding mean value. The COV was used to eliminate the possible positive correlation between magnitude and within– or between–subject variation (ref).

III. RESULTS

In this section, the results of duration, pitch and formant frequencies measurements are presented, together with the results of temporal and spectral variability measurements. Implications for the degree of maturation of neuromuscular control for speech production in children speakers are discussed.

A. Duration and temporal variability

1. Vowel

In Fig. 2(a), mean vowel duration averaged across all vowels and subjects in each age group is shown with between–subject variations (i.e., standard deviations). In Fig. 2(b), the mean duration of individual vowel averaged across subjects in each age group is shown. In Fig. 3, within– and between–subject COV are shown as a function of age. Since a repeated measures, two–way ANOVA indicates no significant gender effect in vowel duration ($F(1,477)= 0.12$, $p= 0.73$), averaged values across gender are shown in the figures. Age effect is significant ($F(14,464)= 16.99$, $p= 0.00$).

As can be observed from Fig. 2(a), five and six year–old children show significantly longer durations than older age groups. There is significant drop in vowel duration from age five to age seven ($p < 0.05$) and no significant change from age seven to adult. On average, vowel duration may reach adult range around age eight or nine. It is also observed from Fig. 2(b) that even children of age five exhibit adult–like pattern of vowel–dependent duration. This suggests that the ability of vowel–specific, or vowel–inherent, duration control is acquired before age five. However, their relative timing control for duration among vowels in a given context seems not well established yet. Children of age five and seven show a tendency of overshoot, or sometimes possible undershoot, of

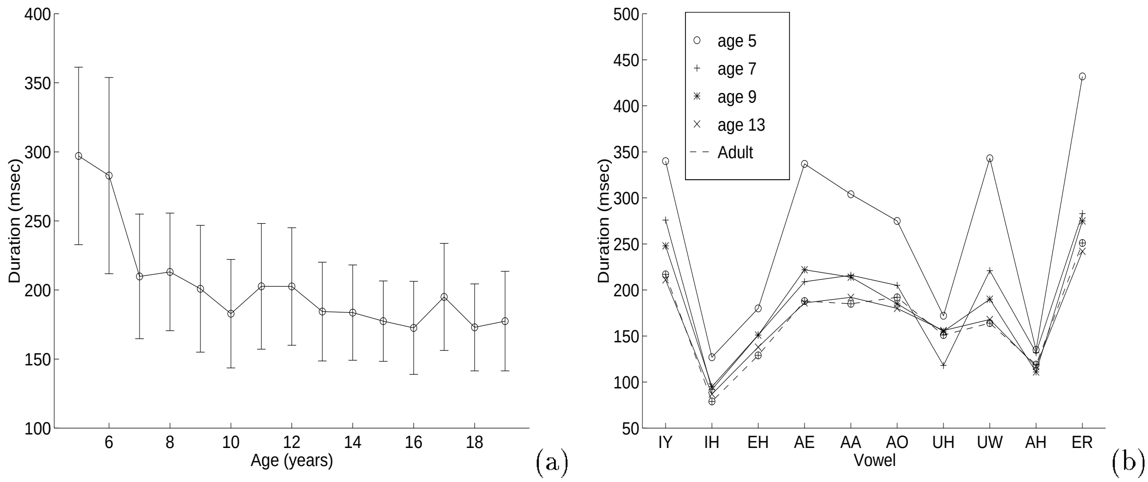


Figure 2: (a) Mean vowel duration averaged across all vowels and subjects in each age group. (b) Individual mean vowel duration averaged across all subjects in each age group.

vowel duration, suggesting that the dynamic range of vowel duration is larger for young children than for adults.

As shown in Fig. 3, the within–subject variability of vowel duration decreases with age and reaches adult range around age eleven or twelve, several years later than duration. A recent study also reports a similar relation between duration and variability in children's speech (Smith, 1992). It is noted that no clear age–dependent trend is observed in the between–subject variability.

2. /s/ and sentence

Duration and within–subject COV are shown in Fig. 4(a) and (b) for /s/ and in Fig. 4(c) and (d) for sentence, as a function of age. As children of age five and six produced target words only, the duration of /s/ is shown beginning from age seven. As gender effect is not so significant for both cases ($F(1,430)= 4.676$, $p= 0.031$ for /s/; $F(1,483)= 0.187$, $p= 0.665$ for sentence), averaged durations across gender are shown in the figures. Age effects are significant ($F(12,419)= 8.279$, $p= 0.000$ for /s/; $F(14,470)= 23.113$, $p= 0.000$ for sentence).

In contrast to the case of vowel, duration and variability of /s/ and sentence productions indicate that both magnitude and variability reach adult levels at approximately the same time, around age twelve or thirteen, for both cases. As also can be observed from Fig. 4(b) and (d), the within–subject variability of both /s/ and, especially, sentence productions substantially drops from age eight to age eleven where the temporal variability almost reaches adult level. This suggests that a substantial improvement of timing control for coordination of speech articulator is achieved at this age range. It is also interesting to observe that on average, teenagers around age fourteen or fifteen show shorter duration than any other age groups, for both /s/ and for sentence productions. About 40% reduction of sentence duration occurs from age eight to age forteen.

An interesting observation in the sentence duration (Fig. 4(c)) is that children of age five and

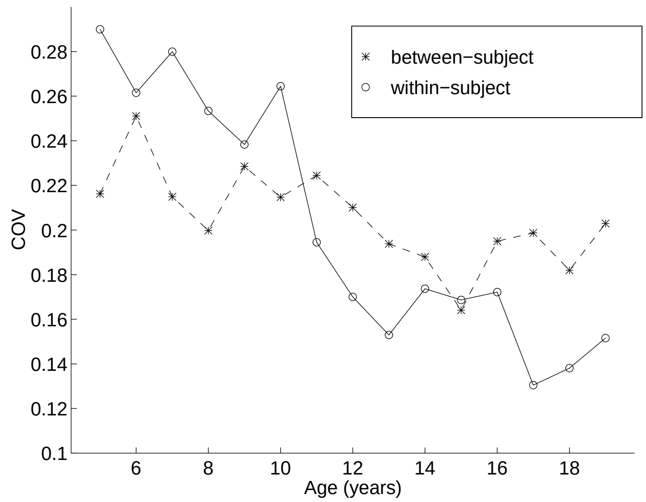


Figure 3: Within–subject and between–subject coefficient of variations (COV) of vowel produtcion.

six exhibit shorter duration than children of age seven and eight. The main reason is found to be that children of ages five and six do not faithfully produce sentences. For instance, they frequently skip some portion of sounds in sentences (e.g., /k/ in "Chuck seems –" and [the] in "– after the race."). In contrast, children of ages seven and eight faithfully produce all sounds in a given target sentence. The outcome is that, irrespective of speech eliciting methods (imitate vs. read), the averaged duration of all five sentence productions is longer for children of ages seven and eight [2,447 msec in reading (83% of total number of utterances recorded) and 2,472 msec in imitation (17%)] than children of ages five and six [2,355 msec in reading (12%) and 2,366 msec in imitation (88%)]. These statistics also indicate that there is a substantial difference in reading ability between the two age groups.

B. Pitch and variability

The mean pitch of male and female speakers averaged across all vowels and subjects is shown in Fig. 5(a) as a function of age. Vertical bars denote between–subject variations. In Fig. 5(b), a distribution of pitch values in male speakers of ages twelve through fifteen is shown.

In overall, age effect and age–gender interaction are significant ($F(14,487)= 40.187$, $p= 0.000$; and $F(14,488)= 13.002$, $p= 0.000$, respectively). However, up to age twelve, gender effect ($F(1,299)= 2.989$, $p= 0.085$) and age–gender interaction ($F(7,293)= 0.510$, $p= 0.826$) are not significant, suggesting that pitch between male and female speakers of the current data base is not differentiated until around age twelve.

For male speakers, there is a significant F0 drop from age eleven to age thirteen ($p < 0.05$) and there is no significant pitch change after age fifteen. This indicates that on average, pubertal pitch change in male speakers in the current database starts between ages twelve and thirteen, and ends around age fifteen. About one–octave drop of pitch value occurs before and after puberty. The

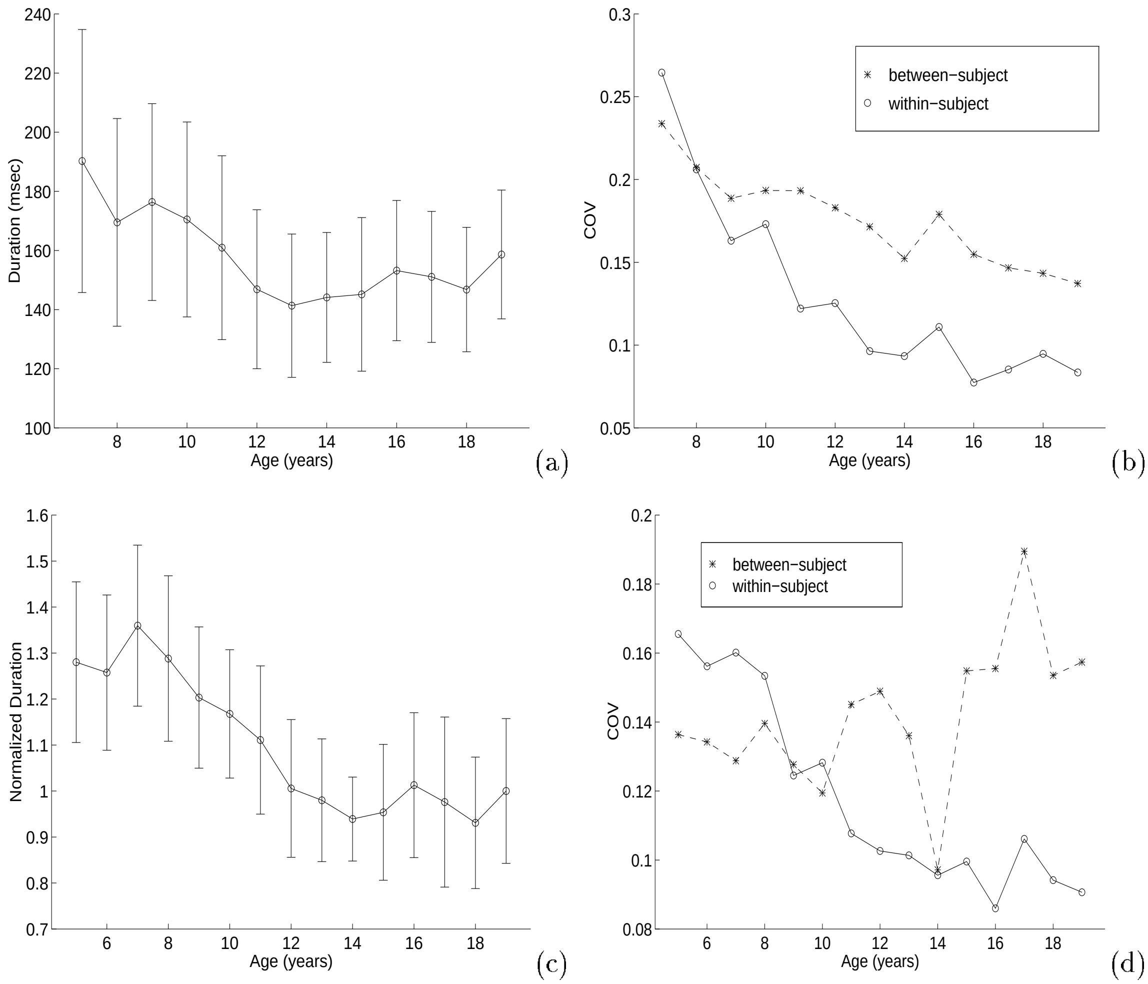
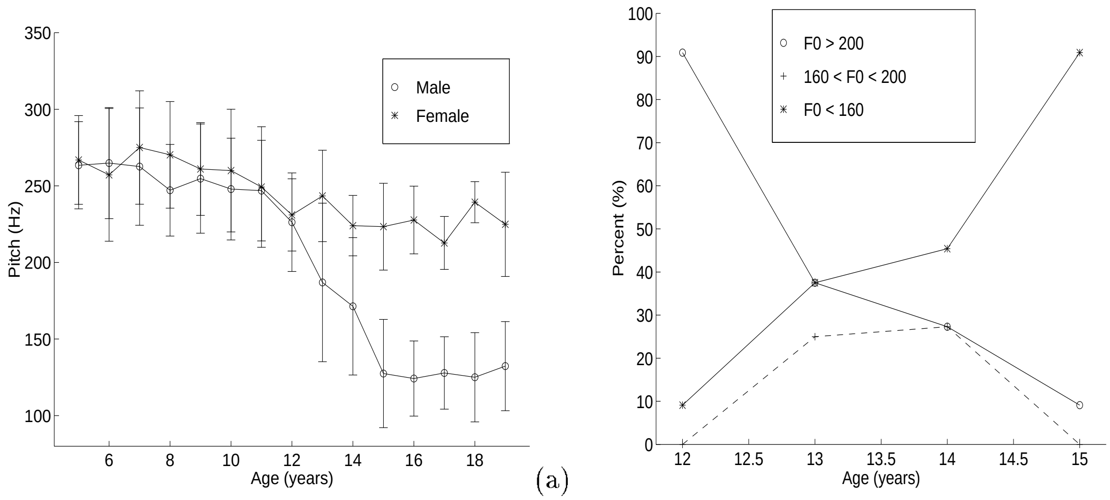


Figure 4: (a) Duration of /s/. (b) Within–subject COV of /s/ production. (c) Normalized duration of sentence. (d) Within–subject COV of sentence production.



(a) (b)

Figure 5: (a) Averaged pitch for male and female speakers. Vertical bars denote between–subject variations (i.e., group standard deviations). (b) Percentage distribution of pitch values for male speakers of ages twelve through fifteen. It is clear that pubertal pitch change begins between age twelve and thirteen and ends at around age fifteen for male speakers in the current database.

relatively large between–subject variations at ages thirteen and fourteen suggests that the onset-time of puberty is speaker–dependent. Specifically, as shown in Fig. 5(b), for male speakers of ages twelve through fifteen, the percentage of speakers whose F0 are less than 160 Hz is 9.1% (2 out of 22 subjects) at age twelve, 37.5% (6 out of 16) at age thirteen, 45.5% (5 out of 11) at age forteen, and 90.9% (10 out of 11) at age fifteen. It is also observed that on average, teenagers after puberty show lower pitch values than adults. For female speakers, the pitch drop from age seven to age twelve is significant ($p < 0.05$) and there is no significant pitch change after that age. While the ages at which pubertal pitch change begins and ends are clear in male speakers as group property, it seems not the case in female speakers. This is likely due to the relatively slower growth of the larynx in female speakers.

Within–subject pitch variation and within–subject COV for pitch are shown in Fig. 6(a) and (b), respectively, as a function of age and gender. It is observed from Fig. 6(a) that for male speakers, the within–subject pitch variation markedly decreases at two periods: between ages six and eight and between ages twelve and fifteen, that is, during puberty. After puberty, the within–subject pitch variation increases again with age. For female speakers, a significant drop of the pitch variation occurs between ages ten and forteen. After age forteen, the pitch variation also increase again with age. This analogy of the within–subject pitch variations between male and female speakers may provide useful information to define the the period of pubertal pitch change in female speakers. It is also observed that on average, the within–subject pitch variation is greater in female than in male speakers, although such a difference is washed out in the variability expressed as COV (Fig. 6(b)).

As can be observed from Fig. 7(a) and (b), as early as children of age five exhibit adult–like

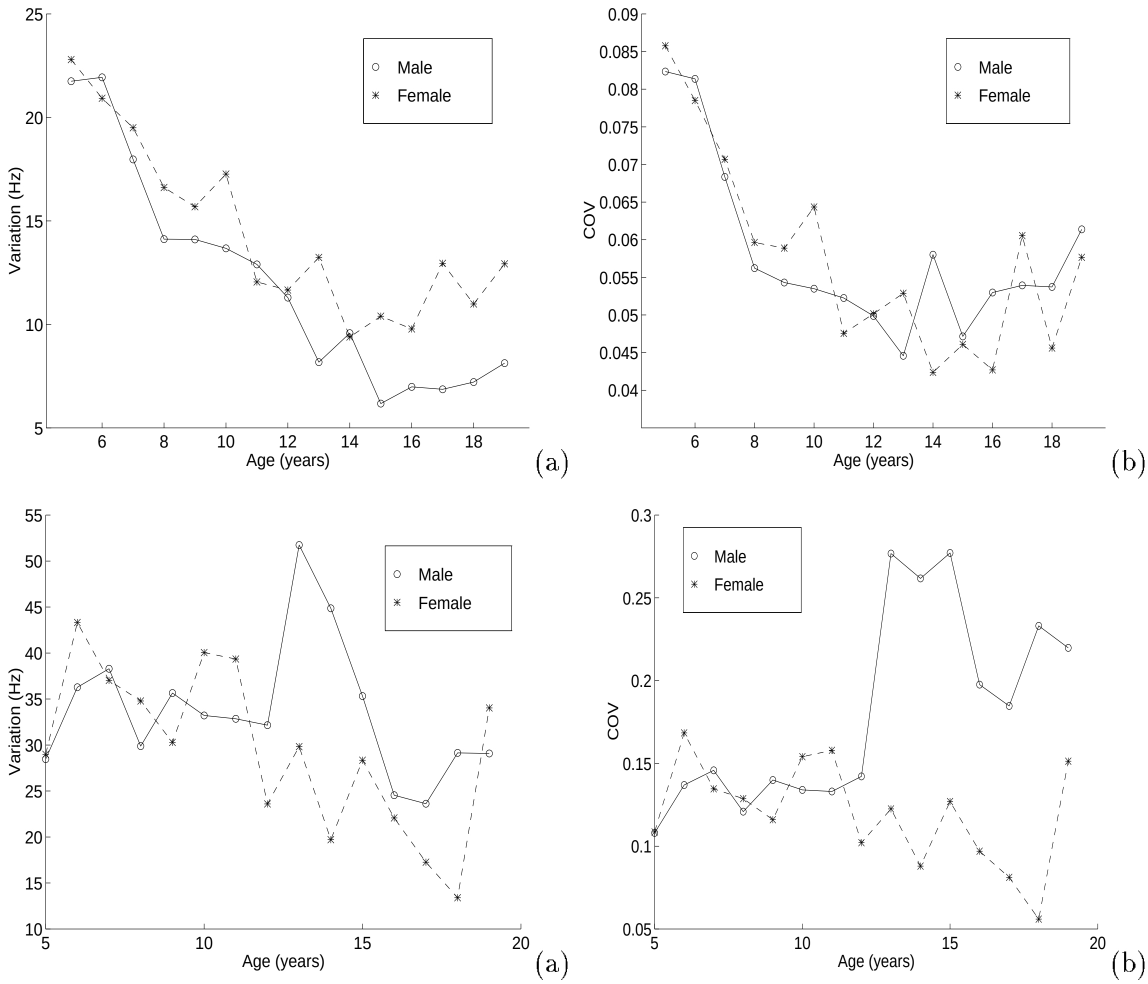
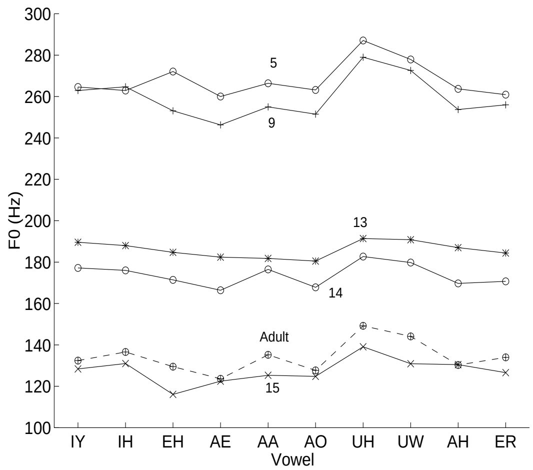


Figure 6: Within– and between–subject variability of pitch. (a) Within–subject variation. (b) Within–subject COV. (c) Between–subject variation. (d) Between–subject COV.



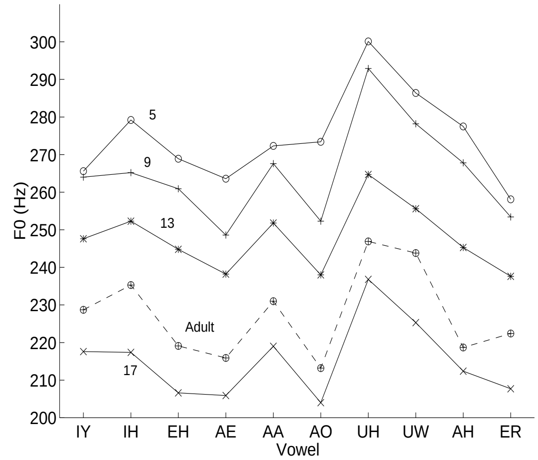


Figure 7: Mean pitch value of individual vowel for (a) Male speakers and (b) Female speakers. A number near each line denotes age.

vowel–dependent pitch patterns. This suggests that the ability of vowel–specific, or vowel–inherent, pitch control is acquired earlier than age five. It is also noted that the pattern of vowel–inherent pitch difference generally agrees with that in Peterson and Barney (1952).

C. Formant and variability

Two–sigma ellipses of several vowels are shown in Fig. 8 in the F1/F2 space. It is observed that the vowel positions produced by 8-years old boys in the current database are slightly compressed or centralized, compared to children's formant data in Peterson and Barney (1959). This centralization of vowel space is most possibly due to the context difference (/h/–vowel vs. /b/–vowel) as well as dialect differences between the speaker population of the two studies (i.e., mid-western vs. pacific eastern). It is also observed that the vowel space as well as formant overlap between vowels are larger in children than in adults. Compariosn of vowel space after an appropriate normalization should indicate a true trned of degree of the overlap among vowels as a function of age.

Scatter plots of mean F1 and F2 of several vowels are shown in Fig. 9(a) for male speakers and in Fig. 9(b) for female speakers. Each point represents mean F1 and F2 averaged across all subjects in a combined every two age groups. For instance, the rightmost circle in /iy/ represents mean F1 and F2 for children of ages five and six, and the leftmost circle for adults. A linear-scaling trend of male formant frequencies as a function of age is clearly observable from Fig. 9(a). Therefore, the current formant data seems to support the notion of "the uniform axial growth of the vocal tract" (ref) for male speakers. Such a linear trend is, however, not clear for female speakers (Fig. 9(b)). This tendecny can be more clearly observed in terms of the formant scaling factor described in the next section.

Formant variability in terms of the within–subject COV is shown in Fig. 10(a) for male speakers and Fig. 10(b) for female speakers. The age–dependent reduction of the formant variability agrees

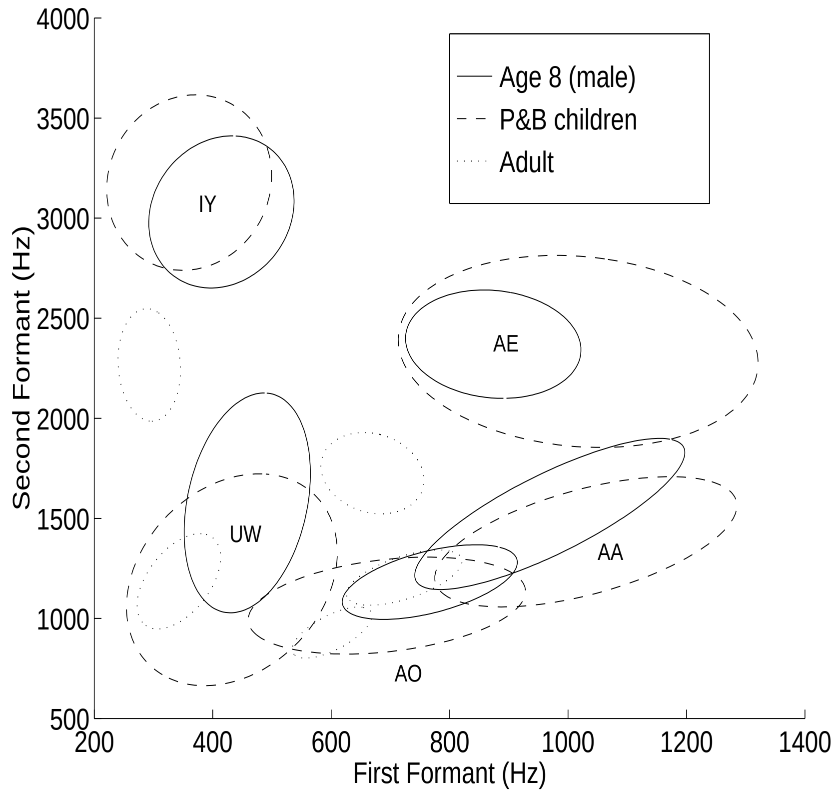


Figure 8: Two–sigma ellipses. Solid and small dotted circles are for children and adults in the current database, respectively, and dashed circle is for children in Peterson and Barney (1959).

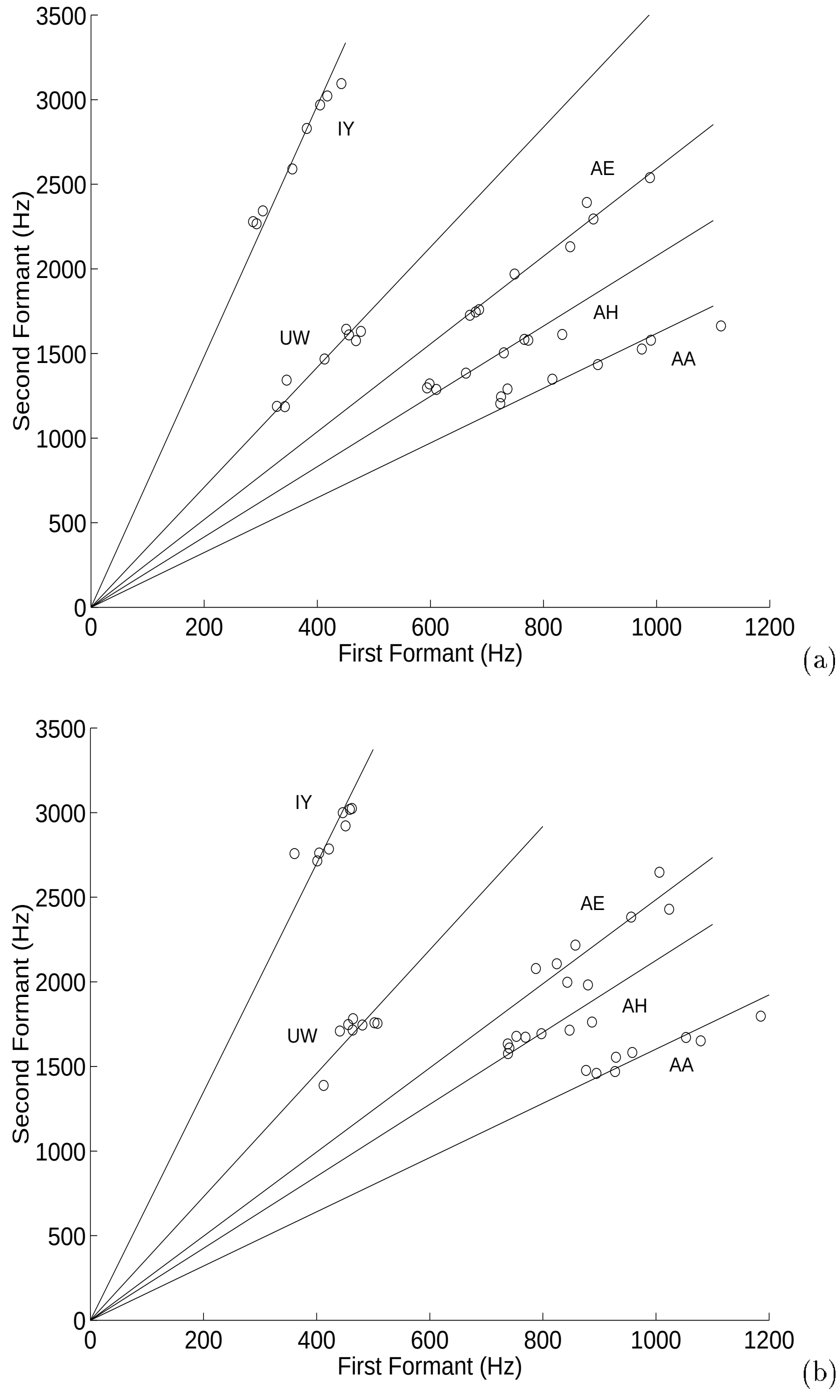
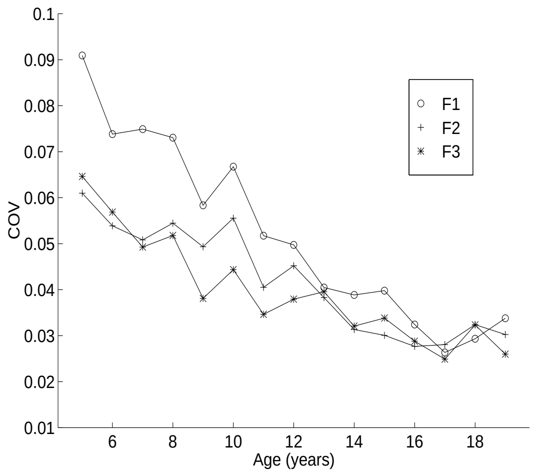


Figure 9: Plot of mean F1 and F2 of vowels /iy/, /ae/, /aa/, /ao/, and /uw/ for (a) Male speakers and (b) Female speakers. The mean F1 and F2 are computed after combining every two age groups into one.



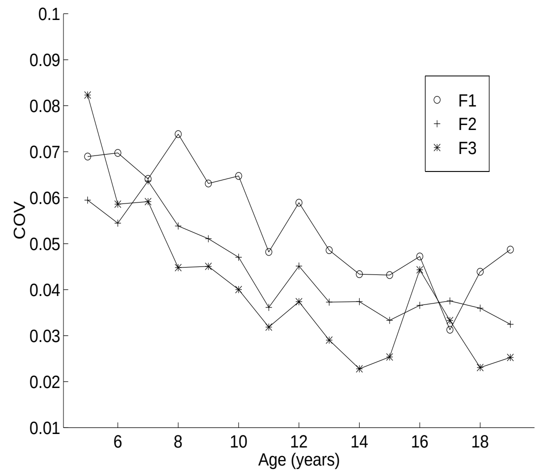


Figure 10: Plot of within–subject COV of F1, F2 and F3 for (a) Male speakers and (b) Female speakers. It is clear that the within–subject formant variability decreases with age.

with the trend shown in Eguchi and Hirsh (1969). It is also observed that the formant variability reaches adult level around age fourteen or fifteen. The variability of the first formant is generally greater than others. No formant–specific trend is observed in the current formant data. That is, the variability of all formants reaches adult range simultaneously. It is noted that in the between–subject COV (not shown), such an age–dependent trend is not observed.

D. Mean formant–scaling factor

In Fig. 11, formant scaling factors scaled by the mean formants of adult male speakers are plotted as a function of age. The scaling factors are computed based on the mean formant values averaged across all vowels and subjects in each age group. Therefore, the major factor which determines the behavior of the mean formant scaling factor is the difference in the vocal tract size as a function of age and gender.

It is observed that differentiation of male and female formant patterns begin around age ten or eleven and the formants become fully distinguishable around age fifteen. Between age ten and fifteen, formant frequencies of male speakers decrease faster and reach much lower values than those of female speakers. This implies that both the total growth and the rate of growth of the vocal tract is greater for male speakers. It is also observed that on average, formant values reach adult range around age fifteen for males and around age fourteen for females.

It is observed that the formant scaling factors of male speakers are approximately the same for F1, F2, F3 and decrease almost linearly as vocal tract grows between ages ten and fifteen. This implies that for male speakers, the acoustic changes resulting from vocal tract growth are uniformly spreaded across vowel formants independent of articulatory differences among vowels. For female speakers, however, each formant evolves differently as a function of age and vocal tract growth. The difference in during and after the development of the front and back cavities between male

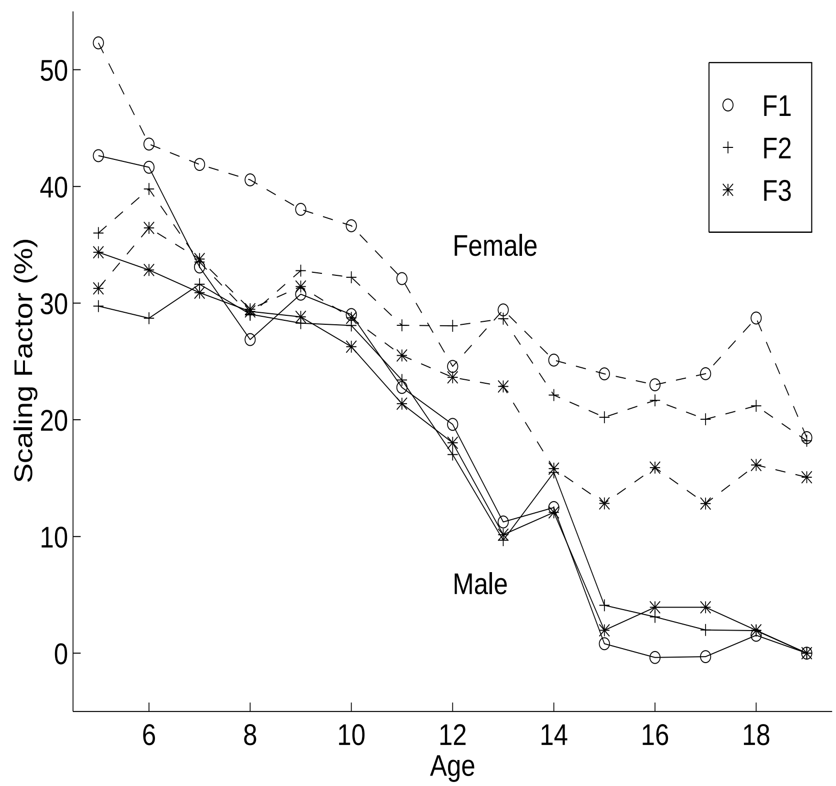


Figure 11: Formant scaling factor for F1 (o), F2 (+), and F3 (*) scaled by adult male formants. They are computed based on mean formants averaged across all vowels and subjects in each age group.

and female speakers may be one of factors that responsible for the differences (ref). Examination of the formant scaling factors of individual vowel, which is not done in the present study, may help to understand the underlying factors for the observation. Another characteristic of female formant patterns is the sudden drop of F1 from age 18 to adult. Physiological and/or socio–psychological factors could be the reason for this phenomenon.

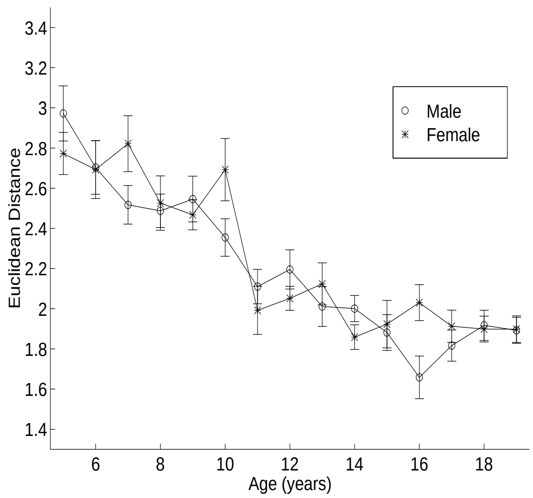
Another interesting observation is the rebound and the drop of formants which occurs between age twelve and forteen in female and between age thirteen and fifteen in male speakers as if they resisted to the acoustic effect of the lengthening vocal tract. The effect is more prominent in male speakers, which is understandable as the vocal tract length grow faster and longer. The similar trend is also observable in the within–subject variability of pitch (see Fig. 6). Exact physiological and/or psychological reasons for these phenomena are not understood yet. However, one reasonable explanation seems that speakers in the middle of puberty are aware of the change of their voice characteristics and they resist to the change, as it is not familiar to themselves yet, in order to maintain as the same voice characteristic as before. The attempt to resist the change of voice quality may be acoustically realized as the instability of pitch control as well as the temporary rebound of formants.

Overall, F3 drops by about 32% from age five to age fifteen for male speakers. This is comparable to the change of vocal tract length (33%) from five year old male children to adults (age 20) predicted in [4]. In [4], however, substantial (12%) vocal tract length growth is predicted to occur between ages fifteen and twenty (i.e., adult), which is not supported by our pitch and formant data. No substantial formant frequency change can be seen in Fig. 11 over age fifteen in male speakers.

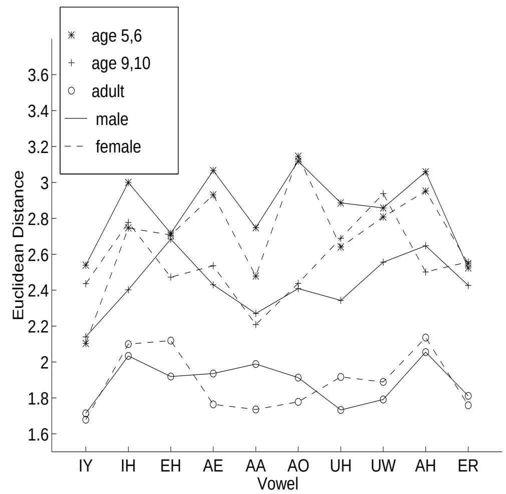
E. Spectral variability

In Fig. 12(a), averaged cesptrum distance between two repetitions of vowels is shown as a function of age and gender. Vertical bar denotes standard error. In Fig. 12(b), it is shown for individual vowel. Clearly, young children exhibit more spectral variation than adults between two repetitions of the same vowels. The tendency disappears around age eleven or twelve and the variability in terms of the whole spectrum reaches adult level around that age. This suggests that young children, especially below age ten, have not fully established their own optimal or stable articulatory targets for vowels. Therefore, it is not surprising at all that the spectral variability and the within–subject variability of vowel duration (see Fig. 3) show a similar trend as a function of age. This is evidence that, in vowel production, the precision of articulatory positioning and the precision of timing control develop in parallel.

The age–dependent reduction of spectral variability also agrees with that of the within–subject formant variability (see Fig. 10). Therefore, the reduction of spectral variability during speech development, in terms of either formant peaks or the whole spectrum, seems to be a genuine phenomenon. It is thus confirmed that the reduction of the within–subject formant variability reported in Eguchi and Hirsh (1969) as well as in the current study is a general tendency that occurs during speech development.



(a)



(b)

Figure 12: Mean cepstral distance between the two repetitions of same vowels is shown for (a) each age group and (b) individual vowel.

In Fig. 13(a), averaged cepstral distance between the initial and the final half of each vowel segment is shown as a function of age and gender. In Fig. 14(b), the averaged cepstral distance is shown for individual vowels. It is clear that even within a vowel utterance, children younger than age ten display greater spectral variability. The tendency is more clearly observable in individual vowels. This is most likely due to excessive and/or abrupt tongue movement during the transition from the vowel to the final consonant /d/, /t/ or /l/. In other words, young children, at least below age seven or eight, have a tendency of more emphasizing the closing tongue movement to the following consonants than older age groups do. If this observation is interpreted as a more distinction between vowel and consonant in intra–syllabic level, this may suggest that young children lack adult–like, smooth coarticulation in VC transition in that the degree of overlap between vowel and consonantal gestures is less in children.

It is also observed that those excessive spectral variations disappear around age eleven or twelve where variability enters adult levels. As shown above, this trend is similar to that of the variability between the repetitions of the same target utterances, suggesting that the precision of articulatory positioning, or the stability of articulatory targets for vowels, develop parallel with the degree of coarticulation skill. In addition, since the within–subject variability of vowel duration as well as both durations of /s/ and sentence and corresponding within–subject variability reach adult levels around that same age (see Fig. 3 and Fig. 4), this suggests again that speech motor control for the precision of articulatory timing and the precision of articulatory positioning follow parallel paths during speech development. Therefore, irrespective of segmental level, the maturity of speech motor control for both timing and articulatory positioning seems to be realized as a single but emergent factor, the degree of coarticulation skill in terms of speed, or agility, and reliability. These results point out that children acquire adult–like coarticulation skill around age twelve.

As can be observed from Fig. 13(a) and (b), female adults generally display greater within–

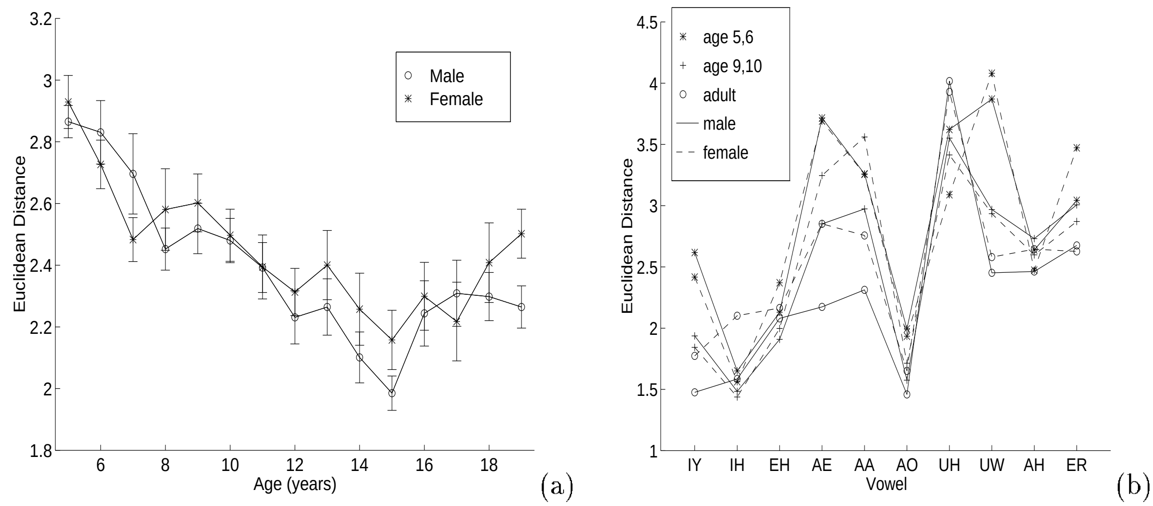


Figure 13: Mean cepstral distance between the first– and second–half segments within vowel is shown for (a) each age group and (b) individual vowel.

vowel spectral variation, or contrast, during VC transition than any other age group older than age nine or ten, indicating more careful or clearer articulation by female adults. It is also observed that teenagers around age fifteen display least within–vowel spectral change among age groups. This is likely due to the faster speaking rate for teenagers (see Sec. A.), which is an example of trade–off between speaking rate and spectral contrast in speech production (ref). In terms of "economy" of speech production, the teenagers seem to best follow the "principles of least efforts (ref)."

IV. DISCUSSION

Many previous studies on speech development in normal children have shown that, in addition to higher pitch and formant values, longer duration and greater temporal and spectral variability characterize children's speech (refs). These characteristics of children's speech emerges due to the smaller vocal tract size and less matured neuromuscular control of speech organs. It is expected that such characteristics disappear as children grow older and adult–like acoustic patterns emerge eventually in terms of both magnitude and variability. The results of the present study are in agreements with those of the previous studies on the characteristics of children's speech, and also well confirm the expected trend of normal speech development, that is, the reduction of magnitude and variability with age in both temporal and spectral domains. In addition, due to the wider and continueous age–range (five through eighteen) and finer age resolution (one–year), it has been possible to obtain more detailed developmental curves of some basic acoustic parameters in both temporal and sepctral domains than the previous studies did. More specific discussion on several findings from this study is described below.

A. Effect of ages on magnitude

This study confirms that the reduction of magnitudes of duration, pitch and formant are a general trend of normal speech development with age. Specifically, vowel duration reaches adult ranges earlier than /s/ and sentence. Since fluent productions of sentences and /s/ in inter–syllabic level (i.e., “I [s]ay”) might require more efficient co–ordination of speech articulators, the result implies that such a coarticulatory skill devleop later after the acquisition of vowel production skill. Therefore, it is likely that duartion in suprasegmental level is governed by a single but collective or emergent factor, the degree of coarticulation skill. Furthermore, duration of sentence almost linearly decreases from age eight to age forteen, which indicates that the speed, or agility, of coarticulation is progressively improved during that age range. In terms of the agility of coarticulation, teenagers around age forteen exhibit peak performance among age groups investigated in this study.

It is noted that there is no clear age–dependent tendency in the between–subject COV of sentence production (Fig. 4(d)). In the case of /s/ production, the Pearson's correlation coefficient between duration and variability is higher for the within–subject variability ($r= 0.51$, $p < 0.00$) than the between–subject variability ($r= 0.32$, $p < 0.00$). This indicates that the within–subject variability is more effective parameter for investigation of developmental change of variability. As shown in Fig. 3, this trend also holds for variability of vowel duration and, in fact, the trend is general tendency for both temporal and spectral parameters investigated in the present study.

imitation (88%)]. In sum, children of ages five and six show irresponsible production of sentences, which might be due to the lack of adequate coarticulation skill to concatenate words as well as due to the difficulty for serial ordering of words in the level of speech motor programming probably due to the lack of vocabulary. For children of age seven or eight, practice in reading at school or home and the expanded vocabulary may explain the faithful aptitude to read all words in a sentence.

B. Effects of age on variability

However, the corresponding within–subject variabilty is substantially decreases from age eight to eleven, suggesting that the precision of timing control is mainly achived at that age raneg.

The latency of within–subject variability stabilization in vowel production may be explained as the time period required to adjust the wider dynamic range of vowel duration.

This study indicates that, irrespective of levels of speech unit, the greater temporal within–subject variability in repetitions of the same utterances is a general trend in young children's speech. There have been arguments that the greater variability in children may be simply due to slower speaking rate (Kent and Forner, 1980), as the precision in timing may be positively correlated with the speaking rate (Klatt, 1974). However, the trend shown in this study seems not due to such an statistical artifact. First of all, such a bias to temporal variability is minimized in this study by taking the ratio of variability to the corresponding mean value. Secondly, this study indicates that teenagers around age around fourteen or fifteen shows faster speaking rate than any other age groups. However, their variability seems almost same, if not greater, when compared to adult. Thirdly, there is evidence that duration of children's speech is still more variable even when duration has been made to comparable to that of adults (Smith et al., 1983). All these evidence point out that the greater temporal variability in children's speech is a genuine phenomenon. Not so well defined specification of segmental duration in the speech motor programming level or above, as indicated by the wider dynamic range of vowel duration, and/or less–refined neuromuscular control of speech organs in the speech motor execution level, as indicated by the less-skilled coarticulation, could be the underlying reasons.

Thus, for male speakers, the pubertal pitch change due to the laryngeal growth and the reduction of the within–subject pitch variability go parallel. It is also observed that the pitch variation is less in teenagers after puberty than in adults. A similar trend also can be observed for female speakers except age difference. That is, the within–subject pitch variation substantially decreases from age ten to age fourteen and then increase again. Although physiological and/or psychological reasons for the reduction of within–subject pitch variation during puberty is not understood yet, the analogy between male and female speakers suggests that pubertal pitch change in female speakers begins between around age ten and eleven, about two–year earlier than male, and ends around age fourteen, about one–year earlier than male. It is observed that the within–subject COV (Fig. 6(b)) also shows the approximately same trend.

Between–subject variation and between–subject COV for pitch are shown in Fig. 6(c) and (d), respectively, as a function of age and gender. No clear age–dependent trend is observed in both figures except the large variations in male speakers at ages thirteen and fourteen due to speaker–dependent pubertal pitch change.

C. Speech development

V. CONCLUSION

In this study, we have measured the magnitude and variability of duration, pitch, and formant frequencies, as well as the variability of whole–spectrum, of children's speech collected from speakers of ages five through eighteen with resolution of one–year of age. Reduction in magnitude and within–subject variability over time are two major indicators of normal speech development. Specifically, when compared to adult, children below age ten exhibit wider dynamic range of vowel duration, longer segmental and suprasegmental durations, higher pitch and formant values, and larger within–subject spectral variability. This trend diminishes around age twelve and, *in both magnitude and variability*, children's speech fully develops to adult levels around age fifteen for male speakers and age fourteen for female speakers. Change of formant patterns in male speakers parallels the vocal tract growth, while for females such a linear trend is not clear. Teenagers around age fifteen differ from both children and adults in that they speak faster, have lower pitch values and exhibit less temporal and spectral variability. It is concluded that the primary factors governing the acoustic patterns during speech development are the anatomical maturation of the speech organs and the maturity of speech motor control in terms of speed, or agility, and precision of coarticulation.

VI. Acknowledgments

The authors wish to thank Jay Wilpon at AT&T for his support throughout the course of this work, and Jim Miller and Rosalie Uchanski at CID for encouragement and helpful discussions.

References

Ananthapadmanabha, T. V. and Fant, G. (**1982**). “Calculation of True Glottal Flow and its Components”, Speech Communication, **1**, 167–184.

J. G. Wilpon and C. N. Jacobsen, ”A study of automatic speech recognition for children and the elderly,” in *Pro. ICASSP*, pp. 349-352, May 1996.

A. Potamianos, S. Narayanan, and S. Lee, ”Automatic speech recognition for children,” in *Proc. EUROSPEECH' 97.*

U. G. Goldstein, ”An articulatory model for the vocal tracts of growing children,” *Ph.D. Thesis*, MIT, 1980.

S. Eguchi and I. J. Hirsh, “Development of speech sounds in children,” in *Acta. Otolaryng.*, Suppl. vol. 257, 1969.

R. D. Kent, “Anatomical and neuromuscular maturation of the speech mechanism: Tutorial,” in *JSHR*, vol. 19, pp. 421–447, 1976.

J. D. Miller, S. Lee, R. M. Uchanski, A. F. Heidbreder, B. B. Richman and J. Tadlock, “Creation of two children's speech databases,” in *Proc. ICASSP*, pp. 849–852, 1996.

A. Ljolie and M. D. Riley, “Automatic segmentation and labeling of speech,” in *Proc. ICASSP*, pp. 473–476, 1991.

B. L. Smith, “Relationships between duration and temporal variability in children's speech,” in *JASA*, vol. 91, pp. 2165–2174, 1992.

G. E. Peterson and H. L. Barney, “Control methods used in a study of the vowels,” in *JASA*, vol. 24, pp. 175–184, 1952.

AGE		aa	ae	ah	ao	eh	er	ih	iy	uh	uw
5	#	26	20	25	28	22	26	23	18	26	22
	F0	264 (32)	262 (35)	272 (34)	260 (37)	266 (33)	263 (35)	287 (42)	277 (35)	263 (31)	260 (35)
	F1	1166 (98)	1010 (117)	824 (129)	837 (78)	835 (101)	620 (70)	636 (74)	467 (67)	656 (59)	477 (45)
	F2	1750 (165)	2534 (163)	1669 (195)	1226 (99)	2475 (216)	1705 (154)	2608 (171)	3071 (145)	1434 (98)	1508 (285)
	F3	3412 (370)	3452 (252)	3596 (300)	3533 (268)	3469 (321)	2502 (353)	3465 (307)	3653 (215)	3588 (353)	3136 (294)
6	#	21	18	18	19	17	15	14	14	18	18
	F0	267 (30)	261 (34)	259 (34)	256 (40)	273 (49)	264 (39)	290 (40)	279 (42)	263 (35)	257 (39)
	F1	1048 (88)	962 (84)	844 (75)	813 (84)	888 (45)	661 (41)	652 (48)	411 (83)	663 (33)	457 (60)
	F2	1554 (117)	2544 (136)	1534 (107)	1147 (117)	2506 (107)	1565 (180)	2728 (141)	3124 (161)	1363 (120)	1659 (270)
	F3	3292 (322)	3316 (131)	3635 (176)	3410 (273)	3417 (262)	2132 (194)	3510 (287)	3728 (216)	3587 (277)	3241 (216)
7	#	19	18	20	17	19	19	16	17	19	14
	F0	266 (42)	268 (42)	258 (37)	253 (38)	265 (44)	260 (39)	281 (61)	281 (40)	264 (40)	260 (43)
	F1	984 (70)	882 (78)	815 (94)	812 (76)	794 (79)	634 (59)	579 (47)	425 (46)	624 (66)	449 (82)
	F2	1536 (90)	2441 (67)	1642 (148)	1223 (145)	2412 (90)	1637 (130)	2602 (111)	3002 (191)	1542 (136)	1700 (236)
	F3	3137 (378)	3344 (147)	3515 (128)	3354 (235)	3440 (214)	2253 (195)	3533 (192)	3618 (219)	3499 (224)	3205 (151)
8	#	41	39	39	40	36	38	40	40	38	39
	F0	257 (32)	251 (29)	249 (30)	242 (35)	247 (32)	241 (30)	270 (40)	265 (39)	248 (38)	245 (31)
	F1	969 (112)	873 (73)	740 (72)	766 (72)	782 (99)	600 (51)	560 (57)	414 (60)	607 (44)	458 (52)
	F2	1522 (186)	2370 (133)	1553 (162)	1181 (92)	2328 (118)	1695 (147)	2568 (130)	3031 (187)	1432 (128)	1577 (271)
	F3	3188 (312)	3254 (199)	3421 (209)	3392 (307)	3409 (282)	2212 (187)	3420 (321)	3626 (285)	3331 (216)	3165 (261)
9	#	37	42	39	36	36	33	33	37	34	34
	F0	262 (45)	264 (39)	253 (33)	246 (32)	255 (37)	251 (31)	279 (46)	272 (45)	253 (40)	256 (41)
	F1	1011 (77)	872 (75)	793 (75)	756 (76)	797 (77)	618 (55)	583 (58)	382 (62)	626 (54)	471 (73)
	F2	1601 (126)	2319 (143)	1529 (137)	1170 (98)	2287 (93)	1577 (142)	2522 (133)	2979 (147)	1472 (124)	1603 (225)
	F3	3245 (234)	3196 (177)	3391 (222)	3375 (233)	3390 (201)	2104 (205)	3466 (210)	3536 (252)	3340 (189)	3198 (224)
10	#	42	41	38	39	37	40	39	43	41	37
	F0	254 (33)	253 (39)	250 (40)	243 (34)	257 (38)	241 (34)	273 (42)	268 (37)	251 (33)	242 (33)
	F1	970 (89)	904 (78)	751 (66)	748 (51)	758 (69)	621 (43)	590 (53)	424 (69)	631 (59)	482 (63)
	F2	1558 (189)	2269 (103)	1628 (147)	1162 (101)	2254 (139)	1644 (149)	2401 (132)	2959 (167)	1501 (174)	1656 (268)
	F3	3148 (202)	3214 (203)	3318 (205)	3198 (318)	3407 (190)	2170 (211)	3434 (195)	3475 (151)	3267 (173)	3049 (314)
11	#	38	38	39	41	39	40	37	42	37	37
	F0	250 (33)	254 (37)	244 (35)	244 (32)	250 (32)	241 (36)	267 (41)	265 (38)	243 (34)	239 (33)
	F1	900 (76)	875 (67)	741 (55)	742 (54)	725 (58)	574 (54)	540 (37)	400 (60)	589 (37)	475 (53)
	F2	1436 (146)	2170 (88)	1525 (129)	1086 (76)	2136 (98)	1539 (121)	2406 (93)	2894 (133)	1479 (179)	1704 (232)
	F3	3075 (257)	3108 (255)	3166 (250)	3182 (275)	3176 (242)	2013 (186)	3290 (219)	3437 (236)	3072 (168)	2939 (180)
12	#	40	37	41	39	37	38	32	36	38	34
	F0	232 (35)	228 (34)	225 (32)	221 (34)	233 (32)	220 (32)	243 (31)	239 (37)	225 (33)	225 (34)
	F1	891 (76)	818 (70)	718 (63)	716 (55)	714 (66)	574 (66)	517 (36)	358 (47)	581 (54)	424 (51)
	F2	1432 (146)	2090 (137)	1484 (149)	1083 (88)	2026 (188)	1530 (124)	2207 (152)	2755 (155)	1450 (157)	1576 (223)
	F3	2930 (207)	3089 (191)	3081 (161)	3039 (233)	3161 (237)	2007 (170)	3191 (181)	3349 (238)	3006 (200)	2805 (208)
13	#	26	25	25	28	29	23	22	25	26	23
	F0	189 (53)	188 (50)	184 (48)	182 (49)	181 (52)	180 (50)	191 (54)	190 (52)	187 (51)	184 (50)
	F1	793 (58)	734 (46)	660 (39)	662 (43)	657 (46)	545 (29)	512 (27)	360 (57)	542 (24)	420 (41)
	F2	1324 (120)	1959 (116)	1388 (164)	1032 (72)	1923 (135)	1392 (122)	2067 (100)	2523 (159)	1396 (120)	1418 (238)
	F3	2664 (202)	2825 (189)	2807 (240)	2831 (261)	2916 (269)	1925 (167)	2860 (174)	3212 (412)	2834 (267)	2672 (285)
14	#	20	20	19	19	18	15	20	21	17	16
	F0	177 (42)	176 (45)	171 (46)	166 (43)	176 (45)	167 (47)	182 (50)	179 (45)	169 (48)	170 (46)
	F1	844 (88)	767 (83)	665 (61)	679 (61)	673 (56)	508 (45)	513 (34)	350 (32)	559 (35)	401 (23)
	F2	1379 (122)	1982 (177)	1376 (71)	1052 (75)	1955 (157)	1484 (118)	2188 (190)	2671 (220)	1379 (86)	1537 (255)
	F3	2679 (163)	2792 (154)	2882 (199)	2829 (184)	2970 (228)	1874 (133)	3040 (222)	3340 (342)	2794 (187)	2645 (154)
15	#	19	18	18	18	15	16	16	20	15	17
	F0	128 (32)	131 (46)	116 (11)	122 (30)	125 (36)	124 (34)	139 (37)	130 (37)	130 (36)	126 (35)
	F1	731 (57)	676 (56)	600 (51)	617 (40)	609 (38)	499 (18)	478 (37)	310 (33)	519 (34)	343 (44)
	F2	1316 (68)	1728 (91)	1385 (134)	976 (47)	1720 (53)	1337 (65)	1992 (152)	2350 (123)	1335 (67)	1316 (209)
	F3	2507 (139)	2573 (111)	2657 (132)	2634 (246)	2648 (105)	1755 (133)	2757 (142)	2964 (268)	2556 (160)	2433 (134)
16	#	17	19	18	17	19	14	16	18	16	17
	F0	126 (25)	131 (26)	122 (24)	120 (24)	126 (30)	118 (20)	134 (29)	131 (25)	122 (26)	123 (21)
	F1	741 (58)	684 (60)	596 (25)	600 (35)	599 (26)	472 (52)	451 (29)	296 (20)	497 (31)	348 (40)
	F2	1261 (78)	1762 (65)	1254 (108)	935 (56)	1766 (105)	1354 (80)	1944 (152)	2334 (177)	1296 (81)	1368 (231)
	F3	2627 (209)	2620 (98)	2682 (185)	2737 (240)	2701 (90)	1754 (88)	2735 (140)	3030 (245)	2611 (114)	2397 (237)
17	#	17	16	16	16	16	10	15	17	15	12
	F0	131 (23)	133 (28)	128 (25)	122 (21)	129 (26)	126 (22)	143 (31)	132 (23)	126 (23)	125 (21)
	F1	713 (43)	685 (63)	585 (20)	612 (32)	581 (25)	491 (38)	457 (28)	289 (26)	514 (28)	322 (17)
	F2	1221 (126)	1759 (83)	1341 (136)	940 (27)	1745 (112)	1284 (42)	1910 (125)	2268 (110)	1348 (100)	1216 (281)
	F3	2637 (127)	2541 (77)	2600 (89)	2689 (187)	2651 (122)	1704 (89)	2724 (154)	3092 (212)	2573 (140)	2468 (242)
18	#	15	17	16	17	16	13	16	17	14	8
	F0	123 (23)	136 (29)	120 (27)	123 (29)	124 (27)	125 (26)	142 (42)	134 (32)	122 (28)	123 (27)
	F1	737 (48)	686 (49)	602 (26)	599 (40)	604 (40)	490 (25)	449 (35)	283 (20)	533 (32)	337 (19)
	F2	1269 (61)	1759 (93)	1252 (70)	881 (52)	1776 (69)	1282 (61)	1955 (117)	2289 (118)	1297 (75)	1144 (169)
	F3	2560 (160)	2560 (96)	2673 (182)	2622 (222)	2633 (89)	1625 (139)	2670 (108)	3050 (238)	2535 (63)	2328 (124)
19+	#	47	47	44	43	46	30	39	49	43	32
	F0	132 (28)	136 (32)	129 (27)	123 (27)	135 (31)	127 (30)	149 (31)	144 (33)	130 (30)	134 (30)
	F1	723 (48)	669 (43)	610 (32)	601 (33)	590 (32)	471 (29)	458 (26)	292 (26)	501 (35)	342 (34)
	F2	1204 (68)	1725 (100)	1288 (119)	929 (62)	1707 (114)	1265 (74)	1851 (110)	2266 (139)	1269 (121)	1185 (117)
	F3	2496 (176)	2532 (151)	2557 (156)	2599 (165)	2549 (134)	1612 (108)	2588 (105)	2930 (184)	2466 (156)	2411 (160)

Table 1: Pitch and formant values for male speakers (# is the number of tokens)

AGE		aa	ae	ah	ao	eh	er	ih	iy	uh	uw
	#	14	16	15	19	16	20	11	13	12	17
	F0	265 (44)	279 (38)	268 (31)	263 (32)	272 (37)	273 (36)	300 (32)	286 (40)	277 (29)	258 (34)
5	F1	1224 (64)	1055 (105)	956 (105)	921 (90)	894 (83)	687 (62)	685 (66)	466 (30)	698 (31)	501 (54)
	F2	1842 (141)	2613 (138)	1772 (108)	1337 (88)	2555 (130)	1707 (137)	2816 (170)	3019 (180)	1376 (65)	1709 (263)
	F3	3435 (387)	3348 (230)	3274 (395)	3675 (250)	3227 (342)	2350 (215)	3526 (289)	3644 (98)	3496 (324)	3332 (161)
	#	25	24	22	25	19	25	23	20	22	20
	F0	271 (45)	264 (39)	259 (44)	248 (43)	265 (53)	259 (41)	282 (41)	280 (58)	258 (38)	252 (40)
6	F1	1163 (53)	972 (121)	839 (52)	844 (76)	828 (58)	654 (52)	643 (66)	433 (71)	670 (61)	512 (67)
	F2	1771 (189)	2671 (183)	1754 (165)	1258 (100)	2661 (115)	1815 (124)	2791 (123)	2986 (151)	1590 (183)	1793 (370)
	F3	3447 (292)	3438 (317)	3759 (233)	3729 (322)	3421 (349)	2458 (381)	3441 (347)	3596 (128)	3665 (381)	3380 (176)
	#	36	28	35	38	21	32	28	35	34	30
	F0	276 (36)	278 (31)	275 (41)	273 (41)	281 (43)	265 (37)	298 (40)	281 (42)	278 (31)	269 (40)
7	F1	1067 (102)	1023 (87)	846 (70)	853 (88)	856 (46)	686 (55)	608 (44)	467 (70)	661 (44)	506 (70)
	F2	1647 (216)	2433 (143)	1723 (162)	1279 (131)	2428 (113)	1740 (184)	2642 (178)	3026 (181)	1566 (110)	1840 (255)
	F3	3493 (361)	3410 (278)	3425 (502)	3569 (336)	3458 (234)	2490 (301)	3482 (283)	3573 (175)	3327 (468)	3316 (255)
	#	14	13	14	16	16	19	11	10	12	14
	F0	274 (42)	280 (35)	264 (31)	268 (40)	273 (31)	261 (33)	303 (50)	294 (49)	268 (37)	260 (39)
8	F1	1108 (64)	1021 (84)	848 (60)	870 (84)	851 (76)	612 (69)	568 (50)	428 (46)	664 (77)	426 (80)
	F2	1660 (57)	2419 (154)	1693 (109)	1274 (73)	2363 (199)	1713 (235)	2674 (192)	2997 (201)	1450 (90)	1539 (165)
	F3	3144 (342)	3271 (282)	3534 (193)	3384 (226)	3263 (352)	2381 (378)	3552 (166)	3604 (164)	3560 (108)	3298 (217)
	#	38	41	38	38	37	37	28	35	38	37
	F0	264 (29)	265 (34)	260 (31)	248 (25)	267 (35)	252 (30)	292 (35)	278 (36)	267 (33)	253 (29)
9	F1	1063 (86)	948 (89)	801 (48)	810 (64)	818 (68)	643 (60)	587 (54)	455 (61)	652 (51)	505 (44)
	F2	1676 (199)	2415 (131)	1658 (144)	1250 (93)	2363 (133)	1757 (163)	2559 (112)	3061 (140)	1506 (107)	1764 (220)
	F3	3284 (275)	3330 (180)	3505 (189)	3387 (248)	3431 (230)	2298 (182)	3516 (205)	3626 (217)	3412 (262)	3247 (195)
	#	25	21	24	24	17	24	22	23	23	22
	F0	261 (35)	267 (45)	259 (45)	255 (48)	263 (41)	258 (44)	275 (50)	273 (44)	257 (38)	258 (42)
10	F1	1037 (95)	970 (68)	791 (79)	822 (92)	860 (62)	612 (72)	609 (51)	472 (45)	636 (51)	496 (54)
	F2	1663 (251)	2318 (140)	1748 (118)	1264 (112)	2306 (92)	1733 (135)	2491 (178)	2969 (134)	1689 (240)	1747 (280)
	F3	3204 (272)	3286 (285)	3431 (227)	3378 (258)	3372 (312)	2264 (258)	3469 (334)	3506 (165)	3307 (270)	3166 (259)
	#	33	34	33	33	31	31	30	28	32	33
	F0	255 (37)	254 (39)	242 (37)	242 (40)	247 (37)	243 (37)	279 (47)	254 (34)	246 (38)	247 (41)
11	F1	980 (83)	878 (84)	765 (58)	791 (59)	775 (50)	638 (50)	590 (49)	467 (57)	637 (39)	475 (38)
	F2	1547 (163)	2219 (138)	1676 (124)	1219 (66)	2245 (109)	1645 (116)	2468 (113)	2971 (96)	1540 (152)	1774 (200)
	F3	3130 (208)	3190 (149)	3282 (165)	3305 (240)	3365 (175)	2167 (148)	3377 (195)	3462 (156)	3207 (139)	3033 (128)
	#	39	35	38	36	35	34	36	37	35	36
	F0	234 (28)	237 (29)	228 (25)	226 (26)	234 (27)	229 (22)	253 (38)	242 (25)	230 (24)	228 (25)
12	F1	939 (108)	836 (128)	742 (100)	761 (56)	712 (87)	620 (47)	537 (73)	439 (52)	591 (58)	452 (42)
	F2	1612 (164)	2215 (188)	1679 (142)	1212 (120)	2189 (170)	1662 (167)	2398 (222)	2884 (182)	1593 (147)	1661 (309)
	F3	3077 (267)	3163 (261)	3287 (221)	3148 (235)	3287 (168)	2247 (398)	3273 (270)	3376 (256)	3148 (233)	2963 (193)
	#	22	23	21	23	21	20	21	22	21	20
	F0	247 (29)	252 (37)	244 (28)	238 (32)	251 (43)	238 (27)	264 (45)	255 (32)	245 (31)	237 (28)
13	F1	959 (71)	824 (60)	779 (62)	753 (49)	770 (49)	627 (37)	592 (42)	426 (62)	654 (32)	490 (51)
	F2	1552 (150)	2187 (134)	1715 (150)	1212 (106)	2199 (96)	1668 (141)	2382 (94)	2861 (77)	1557 (162)	1858 (198)
	F3	3072 (198)	3110 (163)	3245 (148)	3132 (215)	3238 (109)	2161 (187)	3296 (117)	3391 (136)	3146 (179)	2968 (117)
	#	19	19	17	18	18	17	16	18	17	17
	F0	226 (19)	231 (16)	221 (26)	217 (21)	225 (27)	219 (21)	248 (33)	232 (24)	221 (19)	224 (19)
14	F1	893 (88)	824 (77)	756 (55)	746 (58)	736 (50)	627 (57)	573 (59)	415 (28)	630 (44)	433 (35)
	F2	1556 (72)	2010 (100)	1619 (101)	1192 (116)	2032 (114)	1639 (91)	2189 (92)	2693 (130)	1595 (112)	1693 (186)
	F3	2904 (232)	2935 (175)	3019 (173)	2972 (292)	3103 (190)	2115 (181)	3075 (181)	3222 (190)	2966 (194)	2724 (152)
	#	18	19	18	19	18	18	16	19	18	17
	F0	229 (29)	226 (25)	219 (28)	218 (27)	220 (27)	220 (29)	246 (35)	238 (33)	225 (31)	222 (27)
15	F1	900 (48)	851 (86)	738 (42)	767 (39)	744 (36)	594 (38)	564 (51)	378 (27)	620 (53)	434 (22)
	F2	1541 (91)	1942 (99)	1646 (131)	1177 (74)	1971 (118)	1586 (112)	2171 (75)	2653 (153)	1533 (116)	1727 (254)
	F3	2753 (197)	2907 (150)	2931 (157)	2991 (236)	2998 (165)	1949 (109)	3024 (105)	3237 (183)	2827 (149)	2674 (107)
	#	18	20	20	20	21	19	18	19	20	18
	F0	233 (28)	234 (21)	226 (20)	222 (20)	225 (20)	217 (22)	255 (32)	240 (25)	231 (24)	227 (23)
16	F1	851 (56)	835 (68)	737 (44)	749 (43)	735 (47)	602 (43)	541 (46)	423 (52)	613 (48)	447 (35)
	F2	1412 (98)	2050 (100)	1620 (199)	1159 (88)	2003 (143)	1628 (120)	2207 (89)	2776 (147)	1617 (162)	1691 (232)
	F3	2896 (375)	3004 (144)	2991 (146)	2966 (195)	3042 (142)	2056 (126)	3072 (111)	3241 (135)	2927 (155)	2866 (275)
	#	17	17	16	15	13	14	16	15	15	13
	F0	217 (17)	217 (21)	206 (15)	205 (20)	219 (20)	204 (21)	236 (26)	225 (22)	212 (24)	207 (17)
17	F1	922 (82)	845 (72)	735 (37)	751 (41)	728 (42)	552 (45)	558 (58)	390 (63)	623 (28)	424 (22)
	F2	1467 (160)	2007 (106)	1514 (171)	1166 (93)	2010 (73)	1581 (135)	2168 (84)	2717 (84)	1575 (198)	1715 (392)
	F3	2803 (135)	2867 (256)	2890 (160)	2921 (161)	2961 (144)	1995 (123)	3024 (165)	3290 (130)	2801 (146)	2685 (138)
	#	18	17	18	17	19	18	16	16	17	17
	F0	246 (16)	250 (17)	232 (18)	233 (16)	242 (15)	230 (16)	262 (24)	256 (19)	238 (14)	237 (13)
18	F1	932 (47)	914 (45)	741 (33)	777 (45)	754 (58)	619 (53)	587 (52)	418 (37)	605 (35)	480 (34)
	F2	1473 (144)	1955 (110)	1631 (123)	1165 (92)	2014 (117)	1569 (142)	2222 (33)	2801 (46)	1579 (140)	1771 (324)
	F3	2914 (153)	2946 (127)	3027 (104)	3042 (207)	3047 (112)	2058 (134)	3080 (117)	3305 (59)	2924 (138)	2860 (89)
	#	48	46	47	44	45	39	41	46	45	45
	F0	228 (30)	235 (40)	219 (35)	215 (36)	231 (40)	213 (29)	246 (40)	243 (36)	218 (35)	222 (34)
19+	F1	894 (76)	787 (66)	740 (56)	726 (47)	694 (52)	543 (43)	532 (59)	360 (45)	595 (62)	412 (48)
	F2	1459 (124)	2078 (102)	1609 (135)	1079 (89)	2057 (123)	1481 (132)	2183 (111)	2757 (145)	1522 (140)	1388 (248)
	F3	2950 (252)	2916 (145)	2957 (161)	2986 (220)	3005 (139)	1884 (144)	3064 (136)	3291 (200)	2887 (155)	2804 (235)

Table 2: Pitch and formant values for female speakers.