ESTIMATING CHILD AND ADOLESCENT FORMANT FREQUENCY VALUES FROM ADULT DATA

P Martland[‡†], *S P Whiteside*[‡], *S W Beet*[†] *and L Baghai-Ravary*[†]

‡ Department of Speech Science,
† Department of Electronic and Electrical, Engineering,
University of Sheffield, UK.

ABSTRACT

This paper introduces a model being developed for estimating child and adolescent formant frequency values from adult data. The model approximates adult male and female pharyngeal and oral cavity lengths, and scales these along the corresponding male or female growth curve. The second and third formant frequencies are estimated directly from these scaled vocal tract dimensions. Two methods of establishing the first formant from the scaled data are discussed.

Initial results obtained in the scaling of adult data to child values suggest that age, height and gender are all significant when estimating child formant frequency values. Furthermore, averaging of male and female data is found to be inappropriate since the differing growth rates of males and females imply that vocal tract dimensions cannot be linearly related.

1. INTRODUCTION

Different voice qualities are currently being investigated in an attempt to allow users of synthetic voices to change the way they are perceived by others. Gender and emotion are two of the key areas where improvements are being made. The need for better quality female voices being long since established. For rule-based systems, it is generally accepted that a natural adult female voice quality cannot be created by simply scaling rules developed for synthesising adult male voices [1]. Similar restrictions apply if child / adolescent voices are to be created.

Past research has shown that certain vocal tract dimensions are similar when height, weight and age are controlled, but the subjects should also be considered in terms of their gender-related development patterns. For example, the rates of growth are noticeably different for males and females, and those differences are themselves different for different ages (figure 1).

Consequently, within sample-sets, similarities in vocal tract dimensions may be observable in averaged data, where gender-dependent data may be markedly different. This may lead to difficulties when using such data to create synthetic voices, since the averaged formant frequencies may be raised or lowered in favour of the dominant gender over a specific age range. This is further complicated by the differing growth rates within each gender, which can be estimated by considering a subjects height for a given age.

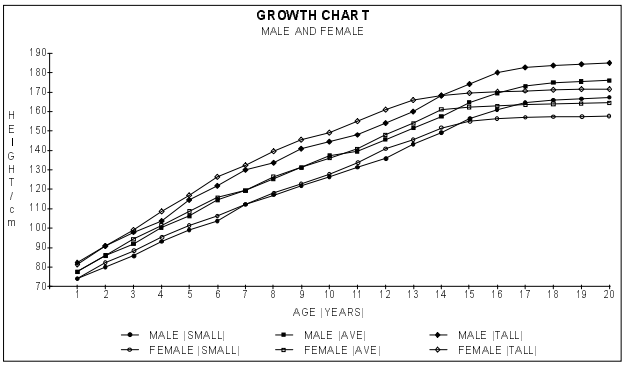

Figure 1. Typical growth curves for males and females over the development years [2].

Previous researchers have examined formant scaling relationships between male/female, male/child, and female/child data [3] employing techniques such as k-factor analysis. Stronger links have been established between the female/child vocal tract relationship when compared to the male/female and male/child relationships. This may imply a more uniform scaling between child and female vocal tract dimensions, but the authors also acknowledge various factors which could influence the data used within the respective analyses.

Vowel quality, and thus formant height may be influenced by dialect in non-homogeneous sample-sets, while variability in anatomical structure may also be influential. Further, if studies are even restricted to single child/adult subjects, one must also assume that articulatory behaviour (for example, the degree of lip rounding) between the two subjects is the same, or at least not significantly influential. By estimating the development of the vocal tract, and by encompassing the predictions of vocal tract theory [3], it may be possible to estimate child/adolescent formant frequency values from adult data, while eliminating the influence of inter-speaker variability. This is the basis of the exponential-quadratic model proposed here.

2. THE PROPOSED MODEL

The purpose of this model is to estimate formant values for child/adolescent ages from adult data. In basic form, the approximations to pharyngeal cavity length (*PL*=35300/2F2) and

oral cavity length (OL=35300/2F3) proposed by Fant [4] are scaled along an 'appropriate' growth curve (determined by the height of the adult subject) until the desired child/adolescent age is reached. The child/adolescent values for F2 and F3 can then be estimated by these scaled vocal tract dimensions.

Although Fant's approximations can yield unrealistically large values in the estimation of vocal tract dimensions, this is not significant since the model is not used to estimate true physiological data: the pharyngeal and oral cavities merely need to be represented in a form appropriate for scaling as if they were single uniform tubes. Further simplifying assumptions have also been made relating to the composition and structure of the pharyngeal and oral cavities.

Composition & Structure of the Pharynx: The pharynx, consisting mainly of skeletal muscular walls approximately 13cm in length for an adult male, lies almost parallel to the cervical region of the spinal column, running from the base of the skull to the 6th cervical vertebra. It is composed of three regions: the nasopharynx, the oropharynx and the laryngopharynx.

The composition of the pharyngeal cavity changes with respect to these three different regions, especially in epithelial composition and mucosal lining. The overall geometry of the pharynx is more funnel shaped than tubular.

Composition & Structure of the Oral Cavity: The oral cavity is composed of two main regions; the oral cavity proper and the vestibule region. The oral cavity proper runs parallel to the anterior hard palate (supported by the maxillary process and palatine bones) and the soft palate, where it junctures with the oropharynx. The internal composition of the oral cavity proper and the vestibule region also vary in epithelial composition, in some places the epithelium being strengthened by keratin.

2.1 Simplifying Assumptions

1. The acoustic correlates of the pharyngeal cavity in isolated vowel production (i.e. not including the nasopharynx) can be represented in terms of the acoustic resonance of a single uniform tube.

2. Similarly, the acoustic correlates of the oral cavity can be represented in terms of the resonance of a single uniform tube.

3. Pharyngeal growth is related to the growth of the cervical region of the axial skeleton, and is thus related to height.

4. Oral tract growth is related to cranial development.

5. The composition and structure of both oral and pharyngeal cavities is consistent through development and adulthood, and therefore any influence via damping/absorption is scaleable.

These assumptions simplify the model, allowing Fant's estimations to be utilised in the scaling procedure and eliminating the non-uniform geometry of the vocal tract. Further, the junction of the oropharynx with the oral cavity is continuous and so difficult to identify. Thus true physical dimensions would be very difficult to establish.

Other simplifying assumptions were also made relating to the limits of scaling: a maximum was set when the growth curve began to level off at 18 years and 16 years of age for the male and female models respectively. The minimum in both cases was set at 2 years, before which growth is far more rapid. The infant also acquires a more adult form of breathing around this age due to the more oblique position of the ribs. Oral tract growth (OLg) was also restricted to 15% in the models over the scaling period since the head at birth is already approximately 75% of its adult size, and rapid growth has already occurred after the first two years. The remaining 85% is therefore a constant (OLc), added to OLg after scaling.

2.2 Relating *PL* and *OLg* to growth curves

Any practical model should take account of both the male and the female growth curves. Therefore the following exponential models for the growth curves were examined:

$$\text{Male:}\quad \ln h = \left(1 - a^{(0.5/\ln(a) - 3.8)}\right) + 4.3$$

$$\text{Female:}\quad \ln h = \left(1 - a^{(0.54/\ln(a) - 3.9)}\right) + 4.3$$

However, both these failed to approximate reliably over the age range of 13-18 years for the male curve (figure 2), and 11-16 years for the female curve. These age ranges are those associated with rapid growth at puberty. Observation of the difference between the estimated height and the true data over these regions for the male and female curves revealed a quadratic relationship (figure 3) which was added as a correction factor to the exponential male and female curves, resulting in realistic approximations to the small growth curves. Further, by using the

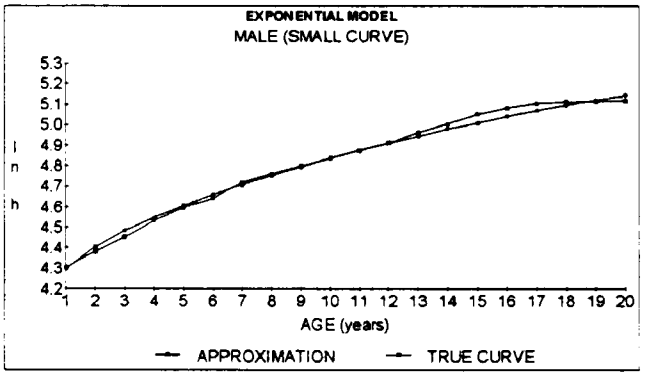

Figure 2. The exponential approximation to the male small growth curve, highlighting the growth spurt associated with puberty.

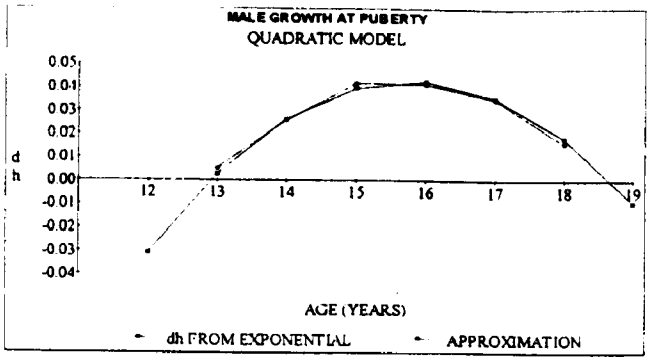

Figure 3. The quadratic approximation for the small male curve used as a correction factor to compensate for the growth spurt at puberty for the male small growth curve.

small growth curves as a reference height (href), a simple linear scaling was employed to accommodate the average and tall growth curves. In the order of small, average and tall curves, the mean errors between the true curves and the approximated curves were 0.92%, 1.23%, and 1.66% for the male curves, and 0.46%, 1.27%, and 3.19% for the female curves respectively.

Fant's approximations for *PL* and *OL* were then used for the final specific male and female scaling models.

SPECIFIC MALE MODEL:

$$PLs = \frac{107h}{F_2 \cdot h_{ref}} \cdot \exp\left(1 - a^{(0.5/\ln(a)-3.8)}\right) \cdot \exp(4.3 + h_c)$$

$$OLgs = \frac{16h}{F_3 \cdot h_{ref}} \exp\left(1 - a^{(0.5/\ln(a)-3.8)}\right) \cdot \exp(4.3 + h_c)$$

$$OLc = 15000 / F_3$$

$$hc = \frac{(a-15.5)^2 - 8.41}{-196} \quad ; \quad 13 \le a \le 18$$

SPECIFIC FEMALE MODEL:

$$PLs = \frac{113h}{F_2 \cdot h_{ref}} \cdot \exp\left(1 - a^{(0.54/\ln(a)-3.9)}\right) \cdot \exp(4.3 + h_c)$$

$$OLgs = \frac{17h}{F_3 \cdot h_{ref}} \cdot \exp\left(1 - a^{(0.54/\ln(a)-3.9)}\right) \cdot \exp(4.3 + h_c)$$

$$OLc = 15000 / F_3$$

$$hc = \frac{(a-13.8)^2 - 10.98}{-262} \quad ; \quad 11 \le a \le 16$$

In both models, 'a' represents the target scaled age, while 'h' represents the height of the adult being modelled from. If the height is not known, then h/href can be simply replaced by a factor of 1.05 in both models. The correction factor accounting for the rapid growth at puberty is represented by hc.

3. ESTIMATING SCALED FORMANT FREQUENCY VALUES

The scaled F2 and F3 values can be obtained directly by manipulating the *OL* and *PL* approximations in terms of the model.

$$F2s = 35300 / (2PLs)$$

$$F3s = 35300 / \left(2(OLgs + OLc)\right)$$

However, the first formant, F1s, cannot be directly established from the results. Thus two methods were investigated. The first method involved estimating F1 in terms of a single tube resonator, where F1 approximates to 35300/4VTL, incorporating the scaled total vocal tract length ($VTLs = Pls + OLgs + OLc$).

$$F1s = \gamma\left(35300/(4VTLs)\right)$$

$$\text{where } \gamma = \left(4(F1_{adult})VTL_{adult}\right)/35300$$

The constant γ here is used to scale the adult F1 value predicted by the simple resonator model to the true value in the adult data. This scale is therefore applied to the estimated values obtained via the exponential-quadratic model.

The second method assumes that the adult F2/F1 ratio can be scaled proportionately along the growth cures,

$$F1s = F2s/\gamma$$

$$\text{where } \gamma = F_{2\ adult}/F_{1\ adult}$$

4. RESULTS AND DISCUSSION

Two speakers from the Eurom 1 database [5] were chosen as subjects, one a male speaker (FA, height 175cm) and the other a female speaker (FC, height 165cm). The first three formant values for the vowels /iy/, /ih/, /eh/, /er/, /aa/, /ao/, /uh/, and /uw/ (using the DARPAbet notation) were extracted from spectrograms of isolated DVD contexts. The F2 and F3 values were then scaled downwards at yearly intervals over the growth period for each gender.

Figure 4 shows the values obtained in scaling the close-mid front vowel /eh/ for both speakers. The F1 approximation in this case is based upon the adult F2/F1 ratio (second method).

What is clearly evident from these projected values for child formant frequencies is that there exists an age at which the F2 and F3 parameters effectively transpose, suggestive that before a certain age, the child vowel space will be determined via the F1 and F3 parameters.

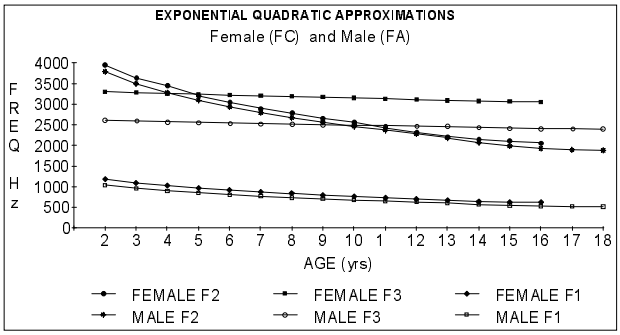

Figure 4. Scaled formant parameters of the vowel /e/ for speakers FA and FC.

However, if we extrapolate the data obtained here towards a younger age than the 2 year minimum of the model, then the formants would compare in value to those observed by Kent and Murray [6] for infant vocalisations (F1=1KHz, F2 = 3KHz, F3 = 5KHz). In this case the pharyngeal cavity would be responsible for the F3 formant, as opposed to F2. This transposition of the F2 and F3 parameters can be attributed to the difference in growth of the pharyngeal cavity and the oral cavity during the development period. the point at which this transposition occurs also appears to be dependant upon gender, and vowel - as well as age.

As a further test, the vowel data extracted for the first three formants has been used within the rule system of a formant synthesiser (OVE III). The vowels above were scaled from the adult male values of speaker FA to the ages of 16, 12, 10, 6, and 4 years. This encompasses ages during and before puberty, as well as child formant values. The formant frequency values were implemented within the definition file, and for improved naturalness, the fundamental frequency values were scaled to an appropriate value (estimated from a plot of F0 against age [7]). For the case of where the F2 value was estimated to be greater than the F3 value, the F3 and F2 positions were transposed in the definition file.

Perceptually, the results of these initial trials of the scaling model were quite convincing, although more in-depth testing is still required. The second F1 model produced more convincing vowel sounds at the younger ages, while little difference could be observed for the 12 and 16 year old voices. Problems were found relating to the default bandwidth of the first formant at the two youngest ages scaled, but the vowel was still audible.

To test whether the vowel space was determined in the scaling by the F3 and F1 parameter for the youngest ages, the F2 and F3 parameters were transposed for the vowel /eh/. However, it was found that the higher formant value was out of the range of the permitted values for the synthesiser in the F2 position. Thus the second strategy just employed removing this value completely, and observing the results. In this case, the vowel quality appeared to have altered, moving more towards shwa, indicating a greater degree of backness.

In conclusion, the model seems to work well, although much more testing is required. Also, the female voice quality needs to be improved before this is scaled and tested. It is known that errors are inherent within the model in particular referring to the assumption of a one to one growth relationship between the upper body and lower limbs. The assumption of a slight but continual head growth is also incorrect, since the head attains full adult size by approximately 12 years of age. However, these effects will be judged more clearly as improvements are made.

The results here, however, do show distinct differences between male and female formant values during the development period. A greater understanding of the relationship between adult and child speech may therefore be achieved if child data is treated in a similar manner to adult data, and is separated by gender.

5. ACKNOWLEDGEMENTS

P.M. is supported by an EPSRC CASE award from Barnsley District General Hospital NHS Trust. L.B.-R. is employed on the EC-TIDE project, "Voices, Attitudes, and Emotions in Speech Synthesis" (VAESS). The authors are grateful to both the Barnsley District General Hospital and the VAESS project for their assistance and co-operation.

6. REFERENCES

1. Karlsson, I. "Female voices in speech synthesis", *Journal of Phonetics, Vol. 19, 1991*, p 113.

2. British Medical Association, *The British Medical Association Family Health Encyclopaedia*, Ed: Smith, T., Dorling Kindersley Publishing, London, 1992.

3. Fant, G. "Non-uniform Vowel Normalisation", *STL-QPSR 2-3, 1975*, p 1-18.

4. Fant, G. "A note on vocal tract size factors and non-uniform F-pattern scalings", *STL-QPSR 4, 1966*, pp. 22-30.

5. Goldsmith, M. *NPL-102 Guide to Eurom 1 Recordings*, Speech Technology Centre, DRSA, National Physical Laboratory.

6. Kent, R. D., Murray, A. D. "Acoustic features of infant vocalic utterances at 3, 6, and 9 months", *JASA, Vol.72(2), 1982*, pp 353-365.

7. Titze, I. R. "Physiological and acoustic differences between male and female voices", *JASA, Vol. 85(4), 1989*, pp 1699-1707.