

Quantification of boys' pubertal status by voice: a trial perceptual study

Martin Ashley
Edge Hill University, UK

Abstract

This paper reports a perceptual test which aimed to quantify, as far as possible, subjective terminology that has been employed by paediatricians and other medical practitioners to describe boys' pubertal status by voice. The study is justified by growing concern with the secular trend in puberty and new national standards for adolescent growth which employ three subjective descriptors of puberty for initial assessment of boys: voice development, testes and penis growth and pubic hair growth. Of these, only the first is non-invasive whilst the second and third are the causes of embarrassment and discomfort to patients. The study shows good prospects for a reliable degree of quantification through voice with prospects for the development of a non-invasive clinical tool for initial GP assessment of precocious and delayed puberty in boys. The tool might also be used in large scale epidemiological studies to track the secular trend.

Introduction

This paper reports a perceptual test which aimed to quantify, as far as possible, subjective terminology that has been employed to describe boys' pubertal status by voice. A key concern is the association between the subjective terminology employed in more general medical literature and the objective measurements that have been made by specialist researchers in the field of voice. New voice-based technologies may present exciting possibilities for research and general medical practice. The study shows that there are good prospects for sparing boys who present with suspected precocious or delayed puberty significant embarrassment or distress, at least during the initial stages. For the gathering of normative data and monitoring of trends the study opens up the possibility of much larger population samples than have hitherto been possible when ethical considerations limit the scope of methods that are invasive of personal privacy. A further important motivation for the study has been the desire to develop a means for interrogation of the secular trend to earlier puberty in boys,(1) (2) relevant to the author's research interest in choral singing.

The previous year (2012) saw the publication of proposed revised standards for adolescent growth by the Royal College of Paediatrics and Child Health (RCPCH). (3) Unlike previous growth charts, these standards attempted to identify points at which the commencement and completion of puberty could be said to be delayed. The height and weight charts are annotated by lines indicating delayed commencement of puberty at age 14:00 and delayed

completion of puberty at age 17:00. The following table in is provided in the chart to assist general medical staff with the recognition of pubertal development:

Figure 1
Identification of boys' puberty phase in RCPCH national standards

Pre-puberty (Tanner Stage 1) <i>All of the following</i>	In Puberty (Tanner Stages 2-3) <i>Any of the following</i>	Completing Puberty (Tanner Stages 4 – 5) <i>Any of the following</i>
High voice	Slight deepening of the voice	Voice fully changed (broken)
No growth of testes or penis	Reddening of the scrotum and growth of the testes	Adult size of testes and penis with adult pubic hair and axillary hair growth
No pubic hair	Early penile enlargement	
	Early pubic or axillary (armpit) hair growth	

Although the Tanner stages are still widely used, matching observations to the Tanner photographs is a subjective process subject to variable observer interpretation.(4) An exploratory study by one of the chart's authors indicated that Tanner based assessment by general medical staff, as opposed to paediatric specialists, was surprisingly unreliable.(5) One possible reason for this is that a high degree of familiarity is needed with the process and insufficient numbers of boys are seen in general practice. Reference to papers published during the 1940s or 1960s show that practices such as asking boys to stand naked on weighing scales whilst researchers made notes about their pubic hair development were once considered perfectly normal and acceptable. Recent discussion of the topic reveals not only that researchers and clinicians now find it increasingly difficult to conduct invasive measurement as a result of contemporary ethical standards,(6) but that boys' embarrassment levels are rising further due to the feminisation of the paediatric profession.(7) Given the levels of distress reported in this blog, the absence of formal research papers addressing this issue more recently than the 1990s(8, 9) is surprising. A number of studies have evaluated the reliability of self-assessment by means of Tanner based drawings in order to overcome such difficulties, but these studies generally report this approach to be disappointingly unreliable.(10-12)

Problems such as these were recently discussed in a paper by Ong et al(13) who, like the present author, were interested in medium to long term trends in pubertal timing. Assessment of the secular trend requires data that predate the establishment of the current Tanner reference standards during the 1970s (14, 15) Ong and his colleagues appear to have placed a high degree of trust in pubertal assessments made in 1961 on the basis of voice. School doctors working within the MRC 1946 National Cohort Study(16) were asked in 1961 to assess the male cohort at the age of fourteen. These doctors were asked to rate the appearance of the genitals as "infantile", "early" or "advanced", to rate pubic hair as "none",

“yes, sparse” or “yes profuse”. They were asked to rate voice-breaking status as “no”, “starting to break” or “completely broken”. Pubertal status as a result of these vocal assessments appears to have been treated as valid and reliable. This seems at least questionable since the meaning of these terms appears not to have been defined. Critically for the subject of the present paper, an assumption appears to have been made that school doctors can make reliable and valid judgements about boys’ voices.

Significantly, similar subjective terminology to describe the boy voice appears to have been employed in 1961 and 2102:

Table 3
Subjective voice descriptors used in 1960 and 2012

Pubertal/genital status	1960 terminology	2012 terminology
Pre-puberty (infantile)	No break	High voice
In-puberty (early)	Starting to break	Slight deepening of the voice
Completing puberty (advanced)	Completely broken	Voice fully changed (broken)

Arguably, judgements of voice such as “high”, “slight deepening” or “starting to break” are at least as subjective as assessments of genital appearance or perhaps even more so. Such subjectivity at the present time is needless as technologies for the assessment and measurement of voice have advanced considerably(17). Moreover, it has been established that there is a close and reliable correspondence between pubertal vocal stage and the key pubertal indicator of testis volume. (18) (19)Correspondence between pubertal vocal stage and Tanner stage is also said to be good in these sources. A more recent study appears to indicate surprisingly high correlations between serum testosterone (ST) levels and voice. Using technology previously unavailable, this study compared speaking fundamental frequency (SF₀) with ST in a sample of forty young adult males (aged 18 – 25). The purpose of the study was to investigate the effect of well-known diurnal variations in ST. Participants self-administered a test protocol involving ST sampling and voice recording at three times of day (09:00, 12:00 and 15:00). Significant negative correlations were found across the sample with SF₀ rising as ST fell (20) Given the non-invasive nature of vocal assessment, further exploration of voice-based methods is justified. Contrary to distress levels reported elsewhere for other methods, the author has found boys to be positively interested in the changes to their voices as a pubertal indicator.(21)

The male voice at puberty

The assumption that the identification of a boy’s voice as “broken” is a straightforward judgement is problematic. Whilst the term “breaking” is in common parlance, this term is rejected almost universally by phoniaticians and vocal coaches. There are good reasons for this. Those concerned with boys’ identity and self-esteem as singers reject the term on the grounds that it implies a personal deficit and sudden loss of musical ability. Studies of what actually happens to the young male voice during adolescence reveal that “break” can refer to a number of different vocal phenomena such as “cracking” (the involuntary flip between two

different registers) or the literal split of the voice into two different halves. (22-30) Voice “break, moreover, is not the unambiguous biological event that is assumed. Boys exercise a significant degree of agency with the regard to the way they speak and sing which makes the objective identification of an actual time of “breaking” surprisingly difficult. (31-35) “Change” less ambiguously refers to the process of vocal metamorphosis which closely parallels the other changes of puberty(18). Whilst many researchers report on this phenomenon, the so-called “eclectic” scheme of John Cooksey is still regarded as the definitive reference.

Table 1
Cooksey’s “eclectic” scheme

Stage	Mean SFo	SFo Range	Tessitura	Full range	Quality
0	259	220-260	D4-C5 (seventh)	A3->F5	Full, rich soprano. The “pinnacle of development”.
1	226	220-247	B3-G4 (sixth)	A3-D5	Breathy, strained upper range; little resonance or “body” in lower range
2	210	196-233	A3-F4 (sixth)	E3-C5	Loss of agility, falsetto emerges, uniquely beautiful and rich if in range
3	186	175-185	F3-D4 (sixth)	D3-A4	Evolution of modal register into baritone range, retention of stage 2 quality
4	151	131-165	D3-A3 (fifth)	A#2-D#4	Light and husky, approximating mid-baritone, difficulties with 4ths and 5ths
5	120	110-139	C3-B3 (seventh)	<A2-D4	Body, resonance and power increase, agility recovered, adult qualities emerge

(Compiled from Cooksey, 1992)(36)

The abbreviation SF₀ refers to “Speaking Fundamental Frequency” (commonly abbreviated as SF₀ or SFF). By this is meant simply the frequency of vibration of the lowest harmonic of the sound heard. In the case of the human voice it is the actual frequency with which the vocal folds collide. In recent years, there has been a proliferation of voice assessment software which allows accurate tracking of the SF₀ during running speech or assessment tasks such as backward counting. Electroglottography allows assessment to be made through direct measurement of the vocal fold motion or mucosal waveform. Two small electrodes placed on the alai of the thyroid cartilage pass a high frequency current through the larynx which is modulated by the vocal folds. Although the Laryngograph remains the industry standard clinical tool,(17) inexpensive alternatives such as the Voce Vista system are now giving good results in scientifically orientated voice tuition.(37)

“Tessitura” refers to the comfortable singing range for most boys and Cooksey’s work demonstrates clearly how this descends in steps closely associated with pubertal stage. Cooksey also provides subjective descriptors of vocal quality which are of considerable significance in the assessment of pubertal vocal status, since there is a wide degree of overlap in SF₀ between stages. Adaptation of the Visual Analogue Scale (VAS), commonly used for patient reporting of pain, has been found to give good results in boys’ self-identification of

vocal change. (22) Boys recognised with some consistency not only a deepening of the voice but a change in its functioning. The tendency to “crack” or flip between different registers during the steepest phase of the growth spurt was clearly recognised, but boys completing the VAS also showed surprising sensitivity to earlier stages of change.

Particularly promising results, however, were found by Fuchs and colleagues who employed an algorithm which processed the key vocal parameters of jitter, shimmer and additive noise independently of the SF₀.(25) The resultant Goettingen Hoarseness Diagram (GHD) could identify vocal pathology imperceptible to the ear whilst separate analysis of jitter and shimmer could reliably predict the key pubertal progression point from early to high voice mutation six months before this could be perceived purely aurally. In summary, the progress that has been made in computerised voice assessment and studies that relate voice to other pubertal markers has been more than sufficient to render Tanner’s views on the unreliability of voice measurement obsolete.

The perceptual test

The investigation now reported set out to answer the following questions:

- What do the terms “high”, “slightly deepened” and “fully changed” mean to general medical practitioners and are they quantifiable?
- What levels of correspondence exist between subjective impressions of boys’ voices and objective measurements of pitch?
- What levels of inter-observer reliability exist with regard to judgement of these qualities in boys’ voices?
- Are there prospects for improved clinical practice with regard to puberty assessment through voice measurement technology?

Method

Two focus groups were recruited for a perceptual test. The first consisted of thirteen independent school music directors with significant experience of choral work with pubertal boys. The second consisted of nine medical professionals including general practitioners and paediatric specialists. The test employed ten voice samples drawn from a large digital audio data base assembled by the author. Included in the database were recordings of boys aged between 9 and 15 reading the phonetic speech passage *Arthur the Rat*.

Ten recordings of this passage were selected from this database. The highest SF₀ selected was 247 Hz, and the other recordings were selected to be progressively approximately 15Hz lower, the lowest SF₀ being 120Hz. Pre-trialling indicated 15Hz to be close to a threshold of what could be distinguished perceptually in different voices. The voice samples were edited and assembled into two sequences incorporated into an audio file of 6’ 14” total duration. The first sequence presented the voices in random order and the second in descending order from highest to lowest. Approximately 20” of each voice was used, from the beginning of the passage up to the sentence ending “...he would always answer I don’t know.” An adult voice giving brief instructions was edited in between each boy’s voice so that one boy’s voice

would not be heard immediately after another. A five second response interval was also edited in after each voice sample. Playback to the participants was in an acoustically dead room via a high fidelity digital audio system.

The test used the language of the RPCCH growth charts. During the first sequence the participants were asked to indicate their response by ticking one of the following statements:

- This is a “high voice”
- This is a “slightly deepened voice”
- This is a “broken” voice

They were also asked to indicate their degree of confidence in this judgement as either

- Definite
- Borderline

During the second sequence, the participants were asked to listen for the first voice they considered to be “slightly deepened” and the first voice they considered to be “broken”, indicating the number of the voice on their response sheet.

Results

The perceptual judgements of the directors of music and medical staff are summarised in Table 4 below. Columns one and two give details of the ten voices whilst column three shows the random order in which the voices were heard at the first presentation. Columns four and seven show the judgements of the directors of music and medical staff respectively. H refers to the percentage judging the voice “high”, SD to the percentage judging it “slightly deepened” and B the percentage judging it “broken”. Columns five and eight show the percentage of respondents confident about the majority category. For example, in the case of Voice 3, 54% of respondents judged it to be “high” but only 15% of these were confident of their judgement. The remaining 75% felt that it was borderline between “high” and “slightly deepened”.

The 225 and 247 Hz voices were confidently and consistently identified as “high” (i.e. pre-pubertal). There was near-universal agreement that the 195Hz voice was slightly deepened (i.e. in puberty) whilst the 212Hz voice was the one representative of a transitional phase of uncertainty concerning the commencement of puberty. The band of uncertainty for completing puberty was wider. The first perception of “broken” (i.e. completing puberty) occurred at 173Hz but confidence was not achieved until 133Hz, although the lower 120Hz still produced some disagreement.

Table 4
Perceptual judgements of music directors and medical staff

Voice no.	SF0	Randomised as	Judgement DoM	Conf DoM	Mean Staging Point DoM	Judgement Medical	Conf. Medical	Staging Point Medical
1 (TA3)	247	4	H 100%	100%		H 100%	85%	
2 (LM10)	225	8	H 100%	100%		H 92% SD 8%	62%	
3 (LM01)	212	6	H 66% SD 33%	44%	3.78	H 54% SD 46%	15%	3.69
4 (LM19)	195	1	SD 100%	78%		H 8% SD 92%	38%	
5 (TA18)	183	9	H 11% SD 89%	33%		SD 100%	46%	
6 (TA6)	173	3	SD 89% B 11%	22%		SD 69% B 31%	23%	
7 (TA15)	164	10	SD 89% B 11%	33%		H 8% SD 54% B 38%	8%	7.23
8 (LG24)	154	2	SD 33% B 66%	44%	8.39	SD 38% B 62%	46%	
9 (LM58)	133	5	B 100%	100%		SD 8% B 92%	85%	
10 (TA36)	120	7	SD 11% B 89%	33%		SD 8% B 92%	77%	

Discussion

It is perhaps counter-intuitive that respondents found it easier to identify a “slight deepening” than a “complete change” or “break”. A possible explanation is that it takes some time for the vocal tract fully to mature after the completion of puberty. The voices of 15 – 18 year olds typically lack the full harmonic spectrum of unchanged or fully mature voices.(38) However, the identification of slight deepening with around the 200Hz mark accords well with the work of Pedersen who used electroglottography to compare the SF0 of 48 boys 8.7 – 19.5 reading a text with other parameters including height, pubic hair, testes volume and testosterone concentration. She found that an SF0 of 200Hz represented a potentially meaningful cut-off point in which boys with an SF0 of <200Hz could, with some reliability, be said to be “in puberty” according to the criterion of a serum testosterone concentration of >10nmols/l. She concluded that the process merited further investigation and development.(39)

The Fuchs et al study supports the contention that pubertal onset can be reliably identified by voice. Whilst the present study has shown a good level of agreement concerning the perception of “slightly deepening”, the Fuchs study holds the promise of a computer algorithm capable of bypassing and surpassing perceptual judgement of the vocal event indicative of pubertal onset. Unfortunately, the Fuchs study did not demonstrate correlation between voice and recognised objective measures such as testis volume or salivary testosterone concentration (ST). As does Pedersen, Fuchs concludes that such further investigation is merited.

The Harries et al study gave clearer correlations between testis volume and voice than testosterone level. (18) However, the more recent study by Evans et al. supports the need for correlational studies of SF₀ and ST. (20) The authors speculated that other studies which had attempted to examine relationships between SF₀ and ST had produced contradictory findings because time of day had not been controlled. The study also established by means of the digit ratio test (comparing the lengths of the second and fourth finger) that pre-natal testosterone levels had little or no effect on adult vocal status. Endocrine events such as puberty in males or menopausal hormone shift in females, however, would seem to have a striking and consistent impact upon SF₀.

Conclusion

This study has shown that subjective terms such as “slightly deepening” applied to the voices of adolescent boys have shared meaning and can be quantified. Pubertal onset seems easier to identify by voice than the completion of puberty at the present level of knowledge but this situation could change as a result of further investigation. At a very simple level, a CD with voice examples could be issued to GPs. Relatively inexpensive technologies to sample voices could become part of the normal equipment available to the GP. Such technologies might be proved through correlational studies of SF₀ and ST. These are both non-intrusive measures and could be carried out on much larger population samples than has been the case for intrusive studies. There are tantalising prospects that inexpensive computer driven technologies might provide a non-intrusive means of identifying whether normal puberty is progressing. The level of accuracy in general practice could be at least as good as Tanner staging and possibly better. Development of such technology is merited by the potential avoidance of embarrassment, distress or reluctance to submit to examination.

References

References

1. Euling S, Herman-Giddens M, Lee P, Selvan S, Juul A, Sorensen T, et al. Examination of US puberty-timing data from 1940 to 1994 for secular trends: Panel findings. *Pediatrics* (supplement article). 2008;121:S172-91.
2. Herman-Giddens M, Steffes J, Harris D, Slora E, Hussey M, Dowshen S, et al. Secondary sexual characteristics in boys: Data from the pediatric research in office settings network. *Pediatrics*. 2012;1058-68.

3. Royal College of Paediatrics and Child Health. Boys UK growth chart 2 - 18 years. 2012.
4. Coleman L, Coleman J. The measurement of puberty: A review. . 2002;25(5):535-50.
5. Butler G. Tanner stages V puberty phases. 2012.
6. Parent A, Teilmann G, Juul A, Skakkebaek J, Bourguignon J. The timing of normal puberty and the age limits of sexual precocity: Variations around the world, secular trends, and changes after migration. *Endocrine Reviews*. 2003;24(5):668-93.
7. Sherman J. Adolescent boys and genital exams: Reducing embarrassment. *Patient Modesty and Privacy Concerns*. 2012(Sunday, 30th September).
8. Heaton C, Marquez J. Patient preferences for physician gender in the male genital/rectal exam. *Family Practice Research Journal*. 1990;10(2):105-15.
9. Kapphahn C, Wilson K, Klein J. Adolescent girls' and boys' preferences for provider gender and confidentiality in their health care. *Journal of Adolescent Health*. 1999;25(2):131-42.
10. Azevedo J, Brasil L, Macedo T, Pedrosa L, Arrais R. Comparison between objective assessment and self-assessment of sexual maturation in children and adolescents. *Jornal de Pediatria*. 2009;85(2):135-42.
11. Hergenroeder A, Hill R, Wong W, Sangi-Haghpeykar H, Taylor W. Validity of self-assessment of pubertal maturation in african american and european american adolescents. *Journal of Adolescent Health*. 2001;28:197-203.
12. Duke P, Litt I, Gross R. Adolescents' self-assessment of sexual maturation. *Pediatrics*. 1980;66(6):918-20.
13. Ong K, Bann D, Wills A, Ward K, Adams J, Hardy R, et al. Timing of voice breaking in males associated with growth and weight gain across the life course. *Endocrine Research*. 2012;97(8):2844-52.
14. Tanner J, Whitehouse R. Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Archives of Disease in Childhood*. 1976;51:170-9.
15. Marshall W. Growth and sexual maturation in normal puberty. *Journal of Clinical Endocrinology and Metabolism*. 1975;4:3-25.
16. Wadsworth M, Kuh D, Richards M, Hardy R. Cohort profile: The 1946 national birth cohort (MRC national survey of health and development). *International Journal of Epidemiology*. 2006;35:49-54.
17. Abberton E, Fourcin A. Electrolaryngography. In: M. Ball CC, editor. *Instrumental Clinical Phonetics*. London: Whurr; 2008.

18. Harries M, Walker J, Williams D, Hawkins S, Hughes A. Changes in the male voice at puberty. *British Medical Journal*. 1997;77(5).
19. Pedersen M. Normal development of voice in children: Advances in evidence-based standards. Heidelberg: Springer; 2008.
20. Evans S, Neave N, Wakelin D, Hamilton C. The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*. 2008;93:783-8.
21. Ashley M. The english choral tradition and the secular trend in pubertal timing. ms submitted.
22. Willis E, Kenny D. Relationship between weight, speaking fundamental frequency, and the appearance of phonational gaps in the adolescent male changing voice. *Journal of Voice*. 2008;22(4):451-71.
23. Sassi S. Effects of vocal registration training on the vocal range and perceived comfort of the adolescent male singer. 2009.
24. Killian J, Wayman J. A descriptive study of vocal maturation among male adolescent vocalists and instrumentalists. *Journal of Research in Music Education*. 2010;58(1):5-19.
25. Fuchs M, Froehlich M, Hentschel B, Stuermer I, Kruse E, Knauff D. Predicting mutational change in the speaking voice of boys. *Journal of Voice*. 2007;21(2):169-78.
26. Willis E, Kenny D. Variability in speaking fundamental frequency in the adolescent voice. Inaugural international conference on music communication science; 5 - 7 December; Sydney, Australia. www.marcs.uws.edu.au/links/ICoMusic; 2007.
27. Hook S. Vocal agility in the male adolescent changing voice [dissertation]. University of Missouri, Graduate School of Education; 2005.
28. Hesterman P. Going through the adolescent voice change. 2000.
29. Cooksey J. Do adolescent voices 'break' or do they 'transform'? *Voice*. 1993;2(1):15-39.
30. Swanson F. The young male with a changing voice. *Journal of Singing (NATS)*. 1981;38(1):32-4.
31. Williams J. Teach singing to children and young adults. Abingdon: Compton; 2012.
32. Williams J.
The implications of intensive singing training on the vocal health and development of boy choristers in an english cathedral choir. 2010.
33. Ashley M. Technique or testosterone? an empirical report on changes in the longevity of boy singers. *Journal of Singing*. 2010(November/December).
34. Ashley M. The angel enigma: Experienced boy singers' perceptual judgements of changing voices. *Music Education Research*. 2011;13(3):343-53.

35. Trundle D. Changing voices: An approach to adolescent voice training. Milton Keynes: Voicessource; 2005.
36. Cooksey J. Working with the adolescent voice. St Louis MO: Concordia; 1992.
37. Miller D. Resonance in singing: Voice building through acoustic feedback. Princetown NJ: Inside View Press; 2008.
38. Cooksey J. Voice transformation in male adolescents. In: L. Thurman , G. Welch, editors. Bodymind and Voice - foundations of voice education. Iowa City, IA: The VoiceCare Network; 2000. p. 718-38.
39. Pedersen M. Biological development and the normal voice at puberty. 1997.