Fundamental Frequency and Perturbation in Infant Vocalisations

Adele Gregory [1], *Marija Tabain* [2], *Angela Morgan* [3]

[1] Linguistics Program, La Trobe University, Melbourne, Australia
[2] Linguistics Program, La Trobe University, Melbourne, Australia, and
[3] Murdoch Children's Research Institute, Melbourne, Australia

am2gregory@students.latrobe.edu.au, m.tabain@latrobe.edu.au, angela.morgan@mcri.edu.au

Abstract

This paper presents fundamental frequency and frequency perturbation measure (jitter) data collected for one infant over a 23-week longitudinal study. It examines whether there are changes in the fundamental frequency over time, whether they are linked to laryngeal constriction and if there is evidence of instability of the fundamental frequency. Results suggest that changes are apparent in both the fundamental frequency's mean and standard deviation over the course of the study. In addition the percent jitter also fluctuates over time. However the role of laryngeal constriction in these changes is still not fully resolved.

Index Terms: infant language, fundamental frequency, jitter, voice quality, laryngeal constriction.

1. Introduction

This paper examines the change over time of fundamental frequency (f_0) and a related perturbation measure (jitter) as part of a larger study into the laryngeal aspects of infant language acquisition in the early months of life.

1.1. Infant speech in the early months

The process of acquiring language and speech in particular is complicated and multifaceted. Although there has been a great deal of interest in this process most attention has been focused on the very early productions (crying), or on the more recognizable speech-like productions found when infants begin babbling. This has resulted in little work being undertaken on the other vocalisations infants produce during the first six months of life, although that trend is changing [1]. Kent and Murray (1982) produced one of the first and still most influential acoustic studies examining comfort-state vocalisations of infants at 3, 6 and 9 months. In this study they only focused on productions they classified broadly as "speechlike" and at each age, different infants were used. Only average f_0 was published though mention was made that the f_0 often varied within as well as between utterances. The f_0 was reported to average 445Hz for 3-month-olds, 450Hz for 6-month-olds and 415Hz for 9-month-olds. A major observation of their study was the documentation of several types of laryngeal behaviour that were not generally observed in older children or adults with clinically normal speech such as tremor, abrupt or rapid f_0 shifts, breathiness and sub-harmonic components. They suggested that these characteristics might be taken as evidence of instability of laryngeal control [2].

More recently Esling and his colleagues at the University of Victoria, Canada, have begun to examine laryngeal articulations and voice quality in infants [3]. Using an auditory perceptual methodology, they found that cross-linguistically infants begin, over the course of a year, to produce proportionally more sounds that occur in their ambient language, and also appear to do so with the voice qualities of that language. Using a model of the larynx that has been developed by laryngoscopic observations of the adult pharynx and larynx during the production of pharyngeal place and manner articulations (i.e. pharyngeal stop, fricative and approximant) in addition to constricted phonatory modalities (i.e. harsh voice, creaky voice, whisper and whispery voice), they found that six different levels of components are involved in producing laryngeally articulated speech [4]. In particular they found that the aryepiglottic laryngeal sphincter is the primary articulator in the production of so called 'constricted' phonatory modalities. Constriction is thus primarily defined in terms of the degree of sphincteringof the aryepiglottic folds in the larynx [5]. It is this model that has been used in their investigations of phonatory settings and laryngeal segments by infants in the first year of life where harsh voice, creaky voice and whisper are all marked by laryngeal constriction, whereas modal voice, breathy voice and falsetto involve no laryngeal constriction. Bettany (2002) used this model to investigate the pitch in the first six months of life of one Canadian infant and found that the pitch was linked to the phonatory modality in that those modalities evidencing constriction had a lower pitch to those that did not. Over time as the modality proportionally changed so too did the pitch [5].

1.2. Measures of voice control

Fundamental frequency is reflective of the biomechanical characteristics of the vocal folds as they interact with the trans-laryngeal airflow. It also provides insight into other elements of speech such as the mechanical adequacy of laryngeal structures and the precision and extent of laryngeal control. Examining the monthly average fundamental frequency allows investigation of voluntary or controlled changes of vocal fold vibration. In contrast, measuring the monthly average jitter values allows investigation of involuntary fluctuations in vocal fold vibration. The stability of phonatory adjustment is reflected in the amount of short-term variability (perturbation) of the voice signal. Small irregularities in the acoustic wave are considered as normal variation associated with physiologic body function and voice production. Jitter measurements measure vocal fold stability and reflect small, cycle-to-cycle involuntary fluctuations in vocal fold vibration [6]. Jitter is most commonly used in the clinical area of speech pathology to form part of a comprehensive voice exam as voice perturbation levels have been shown to considerably increase in association with laryngeal pathology [7]. More broadly, jitter measures have also been used to investigate voice quality from a non-clinical standpoint, with higher levels of jitter being linked to creaky

ISBN 978-0-9581946-3-1
Accepted after peer review of full paper

and harsh voice qualities. Voice onset and termination typically have much greater frequency perturbation than the steady-state mid-portion of a sustained vowel. Because of this it is advised that the initial and final portions of a vowel should be excluded from jitter analysis. Debate is still occurring as to whether jitter measurements performed on different vowel tokens are comparable. Results from different studies on the effect that different formants have on jitter analysis are either inconclusive or contradictory [8]. Because of this, this study will only make comparisons of jitter values using the same vowel.

This paper will examine whether there are changes in the fundamental frequency over time, whether they are linked to laryngeal constriction and if there is evidence of instability of the fundamental frequency via the jitter measure.

2. Method

2.1. Subject

One female infant from an Australian-English speaking environment was recorded in a natural home setting over a 23 week period during 2009. Recording began at the age of 3 weeks and continued until the infant was aged 26 weeks. She had no clinical history of pre- or postnatal illness or of any speech or hearing disorders. Her mother had reported a healthy pregnancy that had been carried to term (40 weeks gestation).

2.2. Recording Procedure

A Sony DCR-TRV16E digital video recorder with integrated microphone was used to film the infant interacting with her caregivers, or engaged in solitary play. The infant was recorded at a sampling rate of 48 kHz and 16 bit encoding. The camera was positioned on a stationary tripod and directed at the infant at a distance of approximately 1 to 2 metres. This distance was selected so as not to unduly distract the infant or be in the way of her caregivers, whilst ensuring proper capture of the sounds produced by the infant. Care was taken to minimize background noise and to provide consistency in the recording environment across the 23-week timeframe. Due to the young nature of the subject (up to 26 weeks) no elicitation of segments was attempted; instead all sounds produced by the infant during the recording session excluding screams were coded and labeled.

During the six-month longitudinal study, the infant was recorded on average for an hour per week. Due to the nature of the subject involved, a week of recording was occasionally cut short owing to a number of reasons such as sickness or tiredness of either the infant or the caregivers. This did not affect the number of tokens available for the study as some weeks the infant was more vocal than others regardless of the length of time recorded.

Once collected the raw videotapes were transferred digitally onto a personal computer. The audio was then extracted at the same sampling rate. The video was segmented into tokens along with the extracted audio. All segmentation and labeling was conducted by the first author. Since the recordings were performed in a natural context (the infant's home) some sounds needed to be discarded due to background noise, technical difficulties or the ambiguity in determining the source of the sound. In addition, any sounds judged to be produced with an occluded oral cavity were also discarded. Due to this being a part of a larger study into the laryngeal aspects of infant vocalisations, all tokens were broadly transcribed using a simplified IPA script in the phonetic database software EMU [9]. Although phonetic transcriptions of infant vocalisations are not altogether satisfactory they do provide an effective method with which to compare different sounds for further analysis. In the form of a broad transcription as expressed above, there is no expectation that the infant is using the same mechanisms as an adult or attempting to reach adult–like targets in their utterances. Thus a simplified vowel inventory was used where:

[A] – referred to any low vowel, front or back
[u] – referred any non-low back vowel
[i] - referred to any non-low front vowel
[E] - referred to any 'neutral' vowel.

The transcription was aided by visual inspection of the video segments corresponding to the token supplemented by broadband spectrograms and time waveforms. A perceptual voice quality label was also attached following Esling's model of constricted versus non-constricted phonatory settings [3]. The determination between these two categories was auditory-perceptual in nature and was again aided by broadband spectrograms and time waveforms.

2.3. Fundamental Frequency and Jitter Analysis

2.3.1. Fundamental Frequency analysis

Acoustic analysis was conducted using PRAAT [10]. As part of the broader study of laryngeal aspects of infant language acquisition the fundamental frequency had been calculated using a 50ms Hamming window at the temporal midpoint of each voice quality of every voiced token. These data will be presented for a total of 733 [A] vowels.

2.3.2. Jitter analysis

For jitter analysis, a specific set of criteria needed to be adhered to. Each token had to:

- consist of only one vowel and no consonants excluding instances of glottal stop or audible breathing
- have a duration of at least 300ms
- have at least 30 glottal pulses (as calculated by PRAAT)
- have a f_0 standard deviation of less than 50 Hz.

These criteria were used to best approximate adult clinical standards [8]. Using PRAAT, which was scripted to automatically exclude the initial and final 50ms, a standard voice report for each token was obtained. This report listed the mean and standard deviation of pitch, the number of pulses as well as the jitter(local) value. Jitter(local) is the average absolute difference between consecutive periods, divided by the average period [10]. It is also referred to as Jitt by the Multidimensional Voice Program (MDVP), a program used predominantly in speech pathology clinics. Table 1 shows the number of tokens for [A], the only vowel with enough tokens for analysis. Once the data were collated, the results were plotted using R [11].

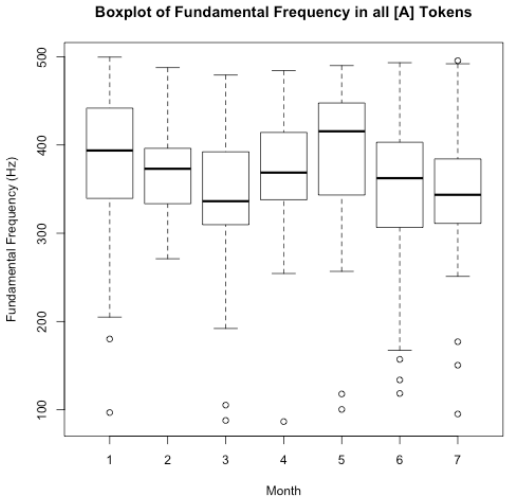
Table 1. *Number of tokens subjected to jitter analysis.*

Month	Number of [A] tokens
1	23
2	12
3	16
4	29
5	11
6	9
7	8

3. Results

Figure 1 presents a box plot of the f_0 for the 733 vowel [A] tokens across 7 months. From Figure 1 it can be seen that during the first three months there is a steady decline in median f_0, however during the fourth month the median climbs and during months five through seven declines again. The inter-quartile range during these last three months is slightly greater than those in previous months and more outliers are present. This might suggest an expansion in control as the infant experiments with different pitches.

Figure 1: *Box plot of f_0 in all [A] tokens.*



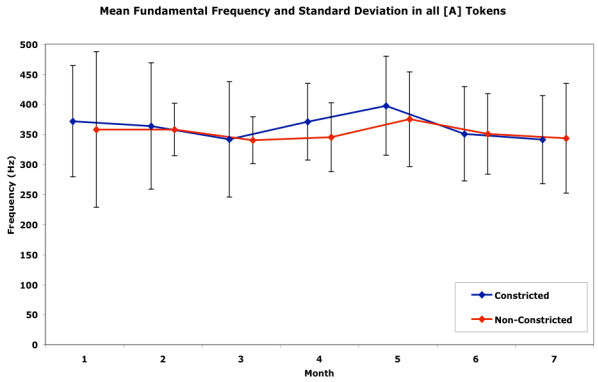
The mean and standard deviation of the f_0 was calculated for each month and then tabulated in Table 2. The mean f_0 for [A] tokens regardless of voice quality was 342 Hz and 349 Hz at three and six months. These values are significantly lower than those reported in Kent and Murray (1982) for the same time period. [2].

Table 2. *Mean fundamental frequency and standard deviation (SD) for all tokens, and separate constricted (con) and non-constricted (non-con) categories. All values in Hz.*

Month	Mean All	SD All	Mean Con	SD Con	Mean Non-con	SD Non-con
1	382	79	372	93	358	129
2	373	57	364	105	358	44
3	343	72	342	96	340	39
4	370	59	371	64	345	57
5	393	79	398	82	375	79
6	349	76	351	78	351	67
7	343	79	341	64	345	57

When these [A] tokens were separated by the presence of laryngeal constriction and graphed against their f_0, there does not seem to be a clear causative link. Figure 2 shows the mean and standard deviation of the constricted and non-constricted [A] tokens.

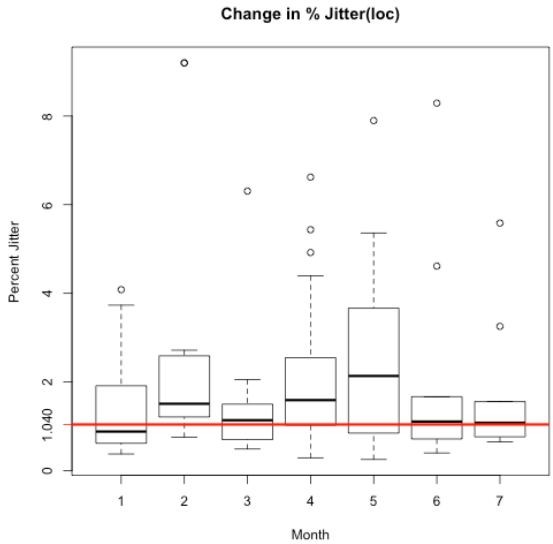
Figure 2: *Mean fundamental frequency and standard deviation for constricted and non-constricted tokens (colour online).*



From this graph it is possible to see that there are no obvious correlations between constriction and f_0 as on average they both follow the same trend and have almost the same mean. The standard deviation of the non-constricted tokens was noticeably smaller between month two and three but by month 4 the two voicing parameters' standard deviations are in close approximation.

Figure 3, below, shows the percent jitter(local) of [A] tokens across 6 months. There are no defined patterns to the levels of jitter as they fluctuate both in median and in inter-quartile range with no real reduction in either over time. A line has been drawn at 1.040%, which is recognized by the MVDP to be the level at a voice can be said to be pathological [11]. This is not to state that this infant's voice is pathological, rather, that at this stage of development, there is still a large amount of variation in vocal fold stability and involuntary fluctuations in vocal fold vibration.

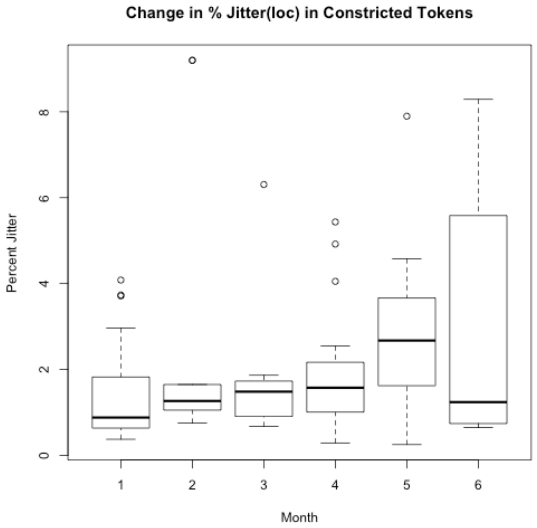
Figure 3: *Box plot of jitter(local) in [A] tokens.*



When these values were investigated for their variance according to constriction, a trend of increasing jitter(local) was found with most of the outliers found in the combined data

being accounted for. Months six and seven were combined due to insufficient data. Constricted tokens accounted for 63% of all tokens analysed for jitter.

Figure 4: *Boxplot of local jitter in [A] constricted tokens.*



4. Discussion & Conclusion

These results show quite clearly that there are changes in the acoustic parameters of infant vocalisations over time. Although the f_0 does not seem to significantly decrease over the period of the study, it does fluctuate in both mean and size of standard deviation. The mean f_0 values determined for this one infant, are significantly lower than those reported in previous studies. This discrepancy may be accounted for by the different voice qualities that were included in this present study. However, due to the small corpus size, it might simply be the case that this infant has a lower f_0. The range of this infant's f_0 is similar to previously reported findings in other infants and children.

When the f_0 was plotted according to the presence or absence of laryngeal constriction, it was apparent that there was very little variability. This is surprising given previously reported findings that f_0 and laryngeal constriction can be correlated due to the laryngeal valves playing a role in both constriction and pitch. It is worth noting here that automated pitch trackers often struggle to analyse constricted voice qualities and that without close manual inspection of the spectrogram and wider harmonic structure as seen in a Fast Fourier Transform, the f_0 can be underestimated. This would lead to an artificial separation between constricted and non-constricted tokens.

The jitter(local) results show very clearly that at this stage in this infant's development a large amount of vocal instability remains. Perceptually this can be heard as roughness and harshness, cues of laryngeal constriction. As such, it would be expected that as the incidence of laryngeal constriction decreases and the voice quality becomes more like the ambient language environment the jitter(local) values will also decrease. A longer study would be needed to verify this claim.

One factor under consideration to explain the results seen here is the anatomical restructuring occurring during the timeframe of this study. During the first three months of life, the anatomy of an infant's vocal and respiratory system are fairly static. However, at approximately three to four months, the larynx begins to descend, the velo-pharyngeal cavity dramatically changes in dimension and the ribs restructure, thus increasing lung capacity [12]. All of these changes cannot be disassociated from the ability of an infant to not only produce but control vocalisations. The abrupt change in the f_0 trend at the three-month mark in light of these anatomical changes certainly raises questions about the exact nature of the link between anatomy, infant vocal production and control. The results presented here are obviously preliminary in nature due to the small corpus size of one. More research is being conducted with additional participants and more methods of measuring laryngeal control and voice quality in order to further clarify these issues. However the large amount of data collected as part of this 23-week study does serve to show that there are changes to the fundamental frequency in both its mean and level of perturbation. The role of laryngeal constriction in these changes is still not fully resolved.

5. Acknowledgements

This work was supported by a La Trobe Postgraduate Award to the first author. We are grateful to Sam Gregory for his support with R and PRAAT and also to the family of our participant for their time.

6. References

[1] Nathani, Suneeti & D. Kimbrough Oller (2001). Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations. *Behavior Research Methods, Instruments, & Computers* **33** 321-333

[2] Kent, Raymond, & Ann Murray (1982). Acoustic features of infant vocalic utterances at 3,6, and 9 months. *Journal of the Acoustical Society of America* **72**(2) 353-364

[3] Benner, Allison, John Esling & Izabelle Grennon (2007) *Infants' Phonetic acquisition of voice quality parameters in the first year of life.* 16th International Congress of Phonetic Sciences Saarbrucken 2007 http://www.icphs2007.de

[4] Edmondson, Jerold & John Esling (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* **23** 157-191

[5] Bettany, Lisa (2002). Range Exploration of Phonation and Pitch in the First Six Months of Life. Master's Thesis. http://web.uvic.ca/ling/students/graduate/Thesis_Lisa_Bettany.pdf

[6] Baken, R.J. & Robert Orlikoff (2000) *Clinical Measurement of Speech and Voice 2nd Ed.* Singular: Australia

[7] Campisi, Paolo, Ted Tewfik, John Manoukian, Melvin Schloss, Elaine Pelland-Blais, Nader Sadeghi (2002). Computer - Assisted Voice Analysis. *Archives of Otolaryngology-Head and Neck Surgery* **128**(2) 156-160

[8] Titze, Ingo, Yoshiyuke Horii & Ronald Scherer (1987). Some technical considerations in voice perturbation measurements. *Journal of Speech and Hearing Research* **30** 252-260

[9] The Emu Speech Database System (2009). http://sourceforge.net/projects/emu/

[10] Boersma, Paul & Weenink, David (2010). *Praat: doing phonetics by computer* [Computer program]. Version 5.1.36, retrieved 18 June 2010 from http://www.praat.org/

[11] R Development Core Team (2003). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. http://www.R-project.org. ISBN:3-900051-003

[12] Kent, R. & H. Vorperian. (1995) *Development of the Craniofacial-Oral-Laryngeal Anatomy* London: Singular Publishing Group, Inc.