Fundamental frequency characteristics of infant vocalisations: a study in voice quality

Adele Gregory [1], Marija Tabain [2]

[1] School of Culture, History and Language, Australian National University, Canberra, Australia
[2] School of Humanities, La Trobe University, Melbourne, Australia

adele.gregory@anu.edu.au, m.tabain@latrobe.edu.au

Abstract

No clear picture exists of the f_0 developmental pattern of typically developing infants. Methodological differences (e.g. type of vocalisations included for analysis) have been found to contribute to this. This paper approaches the f_0 characteristics of infant vocalisations from the perspective of modal and non-modal voice qualities to more fully understand their role in the overall developmental contour. The results presented in this paper support the notion that the f_0 of infant vocalisations provides insight into how an infant learns to exercise vocal control and that voice quality is a useful category through which to investigate these developments.

Index Terms: fundamental frequency, infant, developmental pattern, voice quality

1. Introduction

Previous work conducted on the developmental pattern of typically developing infants' f_0 has been, at times, contradictory. [1], [2] and [3] found no changes in mean f_0 whilst [4] noted a decrease in mean f_0. [3, 918] produced a summary of the available research and concluded (tentatively) that typically developing infants have "high and variable fundamental frequencies in the first year of life in comparison with adults." However no other general trends could be established due to a number of methodological differences in the studies, including:

1. Age of participant,
2. Selection of vocalisations for analysis (e.g. sounds classified as aperiodic, vegetative, squealing, growling),
3. Portion of segment used for f_0 extraction (e.g. nuclei, or nuclei and margins), and
4. How measurements were taken (e.g. visual inspection of waveforms, automatic pitch extraction).

These methodological differences have prevented results being directly compared. Thus a clear understanding of a typically developing infants' f_0 characteristics is yet to be ascertained. This current paper will examine the f_0 characteristics of infant vocalisations by focusing on how the selection of vocalisations for analysis (point 2. above) may give additional insight into a typically developing infant's f_0 characteristics.

It is recognised that vocalisations produced with non-modal voice qualities occupy a large proportion of the sounds an infant produces [5]. These productions include vocalisations with harsh or creaky phonation, those produced with formant structures that are unstable or those with widely fluctuating fundamental frequency. They can also include those that are produced with intermittent voicing or those deemed as vegetative (such as wheezes, sneezes, coughs, hiccups and clicks). Tokens with non-modal voice quality are frequently discarded from analysis in infant developmental studies. In [3, 932] "nearly half of the data had to be discarded because of aperiodicity of the signal, either because syllables were entirely aperiodic (e.g. voiceless sounds) or because they failed to satisfy [their] criterion of having 80% or greater measurable f_0 intervals." Whilst this preference to examine vocalisations with normal phonation is understandable due to it being indicative of emerging linguistic control there is a growing awareness that both modal and non-modal voice uses are important in infants' development of vocal control [5, 553]. This study will therefore revisit infant f_0 from the perspective of modal and individual non-modal voice qualities (creaky, harsh, breathy, loft, whispery voice) in an effort to more fully understand the f_0 characteristics of early infant vocalisations.

2. Methodology

2.1. Recording and segmentation procedure

A Sony DCR-TRV16E digital video recorder with integrated microphone was used to film four infants (3 female, 1 male) interacting with their caregivers or engaged in solitary play over the first six months of life. The infants were recorded at a sampling rate of 48kHz and 16 bit encoding. Due to the young nature of the subjects (up to 26 weeks) no elicitation of vocalisations was attempted; instead all vocalisations spontaneously produced by the infants during a recording session were later coded, unless background noise was present or the infants had occluded vocal cavities. Each vocalisation was broadly transcribed using a simplified IPA script in the phonetic database software EMU. Each participant's vocalisations were also labelled for voice quality according to auditory-perceptual analysis, supplemented by wide-band spectrograms and time waveforms. The qualities considered for analysis were: harsh voice, creaky voice, whispery voice, modal voice, breathy voice, loft whisper and voiceless. Approximately 10% (1140 vocalisations) of the total corpus of two participants were labelled by an independent rater. Inter-rater reliability for this labelling was calculated at a Cohen's Kappa of 0.76 for phonetic segmentation and 0.80 for voice quality.

2.2. Peculiarities of infant data

Infant vocalisations have characteristics that are quite different from those of adult vocalisations in that they have a wider f_0 range, abrupt f_0 transitions and unique energy distribution patterns over frequencies [6]. Software designed to estimate f_0 routinely experience problems determining the f_0 contour within an infant vocalisation. It will mistakenly determine the

f_0 as either double or half what is correct. [6, 205] says that these types of errors are "often considered to be one of the most significant problems of f_0 estimation." Although this can happen for adult data, it occurs more commonly in the data of infants because of the wider f_0 range.

The extensive use of different voice quality modalities also interferes with the f_0 estimation. Segments displaying creaky or harsh voice have voicing discontinuity because of their production. Although the voicing threshold can be lowered in acoustic analysis software programs such as PRAAT to account for these types of segments, excessive use of this setting can adversely affect the reliability of the f_0 tracking by picking up on 'voicing' that is not actually there. These issues make f_0 pattern estimation of infant vocalisations difficult. Because of these factors, it is important to have a robust methodology to deal with difficulties of working with infant vocalisations. For this reason using the software PRAAT, each spectrogram was individually inspected and the f_0 contour corrected when necessary. The f_0 value extracted by the tracker was then compared to the first harmonic of a Fast Fourier Transform (FFT) for verification. This process provided a robust technique for working with a corpus that included so many aperiodic vocalisations.

Rather than examining vocalisations at the level of the syllable, as used in a number of other studies such as [3] this study calculated the f_0 for each voiced segment at the temporal midpoint. This was done to enable the individual influences of the different voice quality modalities to be examined and analysed. Standard deviation was also calculated in a similar manner. A total of 7,517 segments had their f_0 calculated

3. Results

3.1. Longitudinal mean and standard deviation f_0 trends

Figure 1 presents the combined mean and standard deviation for f_0 across the length of the study. In contrast to [4], no linear developmental trend was evident in the infants' f_0 data. This may be accounted for by the comparatively short length of this present study as [4, 1640] suggest that "two or more years of observation would be necessary to obtain a significant tendency for the f_0 decrease." The reported f_0 decrease per 12 months is so small (between 1.9% and 6.1% in their study) that they would be difficult to detect as a tendency. The present data falls more in line with [1], [2] and [3] who found no changes in mean f_0. In this study the mean f_0 decreased until month 3 and after this point it increased again, see Figure 1. Overall mean f_0 for the 6 month study was 367Hz and is similar to those reported previously [2] and [6]. When looked at individually the four children demonstrated considerable variation, ranging from a mean f_0 of 333Hz to 427Hz.

The standard deviation (SD) also shows a similar pattern decreasing in the initial half of the study before increasing again. The SD also varied across children, ranging from 100Hz to 233Hz. Overall the values were generally higher than those previously reported in the literature [3, 918]. This increase in SD may be attributed to the inclusion of a more diverse corpus of infant vocalisations that includes all sounds the infant produced, especially those with non-modal voice quality.

3.2. Longitudinal voice quality trends

Figure 2 presents the mean f_0 for each voice quality plotted by month. Loft has an extremely high f_0 across the entirety of the study. It has a curve that shows quite a steady decrease in mean f_0 from month 1–4 and then a much greater rate of increase from month 4–6.

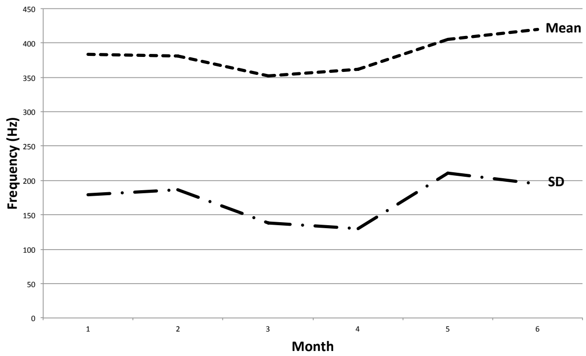

Figure 1: *Combined mean and standard deviation for f_0.*

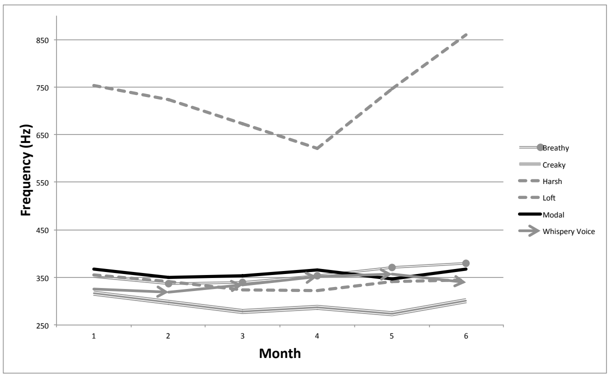

Figure 2: *Longitudinal voice quality mean f_0.*

Figure 3 gives a closer look at the other voice quality modalities. All voice qualities initially experience a decrease in mean f_0 during the first month of the study. For breathy voice this is followed by consistent increases in mean f_0 for the remainder of the study. Whispery voice also experiences increases in mean f_0 for most of the latter part of the study. However, its results must be taken with caution due to the small number of tokens available for analysis (see Table 6). Harsh voice and creaky voice follow a similar pattern to that of loft, though not in the same scale. Decreases in mean f_0 during the first half of the study were again followed by increases in the latter half. Apart from loft, modal voice had the highest mean f_0 for most of the study.

3.3. Variability of voice quality

The coefficient of variation (COV) for f_0 (SD/mean) provides a correction for inter-relatedness, separating variability from absolute values of f_0.[1] When used to examine the variability of voice quality it showed that loft voice had the highest rate of variability, whilst modal had the least. [3] reported that high mean f_0 tends to correspond to high variability in absolute values of f_0. This proves to be the case with loft voice exhibiting the highest amount of variability as well as the highest mean f_0. However modal voice always exhibited the lowest amount of variability, despite the fact it maintained the second high-

[1] The number of tokens can affect the COV. Whispery voice was a small proportion of the data set (see Table 4) therefore its high level of variability should not be given too much importance.

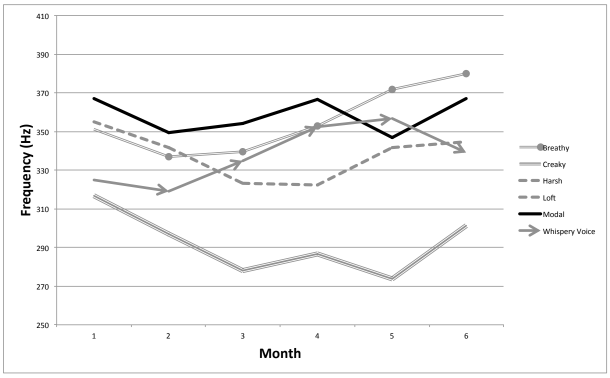

Figure 3: *Longitudinal voice quality mean f_0 (excluding loft).*

est mean f_0. When the data were examined by month, modal voice had the least amount of variability across the entirety of the study. This shows that infants are potentially regulating the amount of f_0 variability that they are producing in different voice qualities. It is important to note that the dominant voice quality in English (modal voice) has a demonstrably higher degree of control being exercised over the use of f_0 and that this is apparent even from the first month of life.

Table 1: *Mean COV for different voice qualities.*

Breathy	Creaky	Harsh	Loft	Modal	Whispery Voice
0.30	0.35	0.42	0.44	0.23	0.40

3.4. Statistical Analysis

A linear mixed effects model was performed which incorporated both random and fixed effects. This analysis was particularly appropriate for spontaneous infant 'speech' because of its flexibility in handling missing values and unmatched numbers of tokens in the individual participants. In addition, mixed effects models offer the advantage of providing insights into the full structure of the data by examining fixed and random effects simultaneously [7]. Analyses were carried out using the R statistical computing software [8]. In the model, the dependent variable was f_0 (transformed into bark for normality). The independent variables in the model included one random-effect factor (subject) and three fixed effect factors (phonetic category, perceptual voice quality and month). Only the results for voice quality and month will be discussed in this paper. A model where there was interaction between perceptual voice quality and month performed significantly better ($\mathcal{X}^2 = 81.094, df = 30, p < 0.001$) than one without interaction. A Tukey post-hoc comparison accounting for interaction was then conducted in order to ascertain significance.

There were some significant effects evident between individual months. Table 2 shows these interactions. These interactions show that there is a statistically significant difference between the lowest f_0 values (months 2, 3 and 4) and the highest f_0 values (months 5 and 6). However across the entirety of the study there are no significant changes in f_0. As such no longitudinal trend of increasing or decreasing f_0 can be ascertained. When considering voice quality there are some significant effects. The f_0 of loft segments and creaky voice segments are significantly ($p < 0.001$) different from all other voice qualities. Creaky voice has a lower f_0 than all other voice qualities, whilst loft has higher f_0 than every other voice quality. Harsh voice also has a significantly different f_0 than a number of the other voice qualities. These interactions are shown in Table 3.

Table 2: *Fundamental frequency interaction by month.*

Month	1	2	3	4	5	6
1						
2					*	***
3					**	***
4					*	***
5		*	**	*		
6		***	***	***		

Statistical significance
$(* = p < 0.05)$, $(** = p < 0.01)$, $(*** = p < 0.001)$

Table 3: *Fundamental frequency interaction by voice quality.*

	Breathy	Creaky	Harsh	Loft	Modal	Whispery Voice
Breathy		***	***	***		
Creaky	***		***	***	***	***
Harsh	***	***		***	***	
Loft	***	***	***		***	***
Modal		***	***	***		
Whispery Voice		***		***		

Statistical significance
$(* = p < 0.05)$, $(** = p < 0.01)$, $(*** = p < 0.001)$

4. Discussion

The results in this study serve to provide further clarification regarding the f_0 trends in typically developing infants. The mean f_0 for the infants of 367Hz was similar in value to a number of previous studies including: [9] and [10]. It is almost identical to that reported by [6]. However it is quite different from that reported by [11] and [12]. The mean f_0 results from these studies were quite high (529Hz and 450Hz respectively). A reason for the lower values found in the present study is the inclusion of all infant vocalisations produced during the recording sessions. This included a large proportion of sounds (see Table 4) produced with extremely low f_0 such as creaky voice segments.

In terms of developmental trends, this present study observed mean f_0 fluctuations month to month and differing patterns between infants. Such individual variation was also observed by [4] and [9]. However longer range trends as seen in [12], [13], and [14] were not in evidence in this data. Instead the results of this study parallel those of the [15] study where no consistent increase or decrease was observed for mean f_0 between 0 months and 6–9 months. The results are also similar to [11], which reported that the mean f_0 decreased between zero and one month and then increased and became stable at 2–4 months. These latter two studies also utilised a methodological approach in which non-modal vocalisations were accepted for analysis and this may have played a role in the comparability of results. The lack of an overall decreasing trend is notable due to the anatomic changes occurring during the timeframe of this study. A decrease in f_0 would be hypothesised due to the lengthening of the vocal tract in both the oral and laryngeal dimensions [16]. However [14] did not find a decrease until after a period of relative stability during the first year. Although a longer and larger scale study would be needed to further clarify the overall longitudinal trends of f_0, this present study does help to reveal f_0 changes over a short time period.

The most significant finding of this current study is the impact of voice quality on infant f_0. Loft, creaky voice and harsh voice are significantly different from other voice qualities based on f_0 alone. It is suggested that the developmental pattern evidenced here for loft vocalisations, played a role in determining the overall contour of the data as seen in Figure 1. While loft vocalisations as a whole only make up a small proportion of each months' productions (see Table 6), the high variability in mean loft f_0 have influenced the overall developmental contour. Harsh and creaky voice also displayed similar developmental patterns. Although their variability was not as great as lofts, they comprised a larger proportion of the data set. Together these three voice qualities (loft, creaky voice and harsh voice) have the largest mean coefficient of variation.

Although infants still have variable f_0 (seen in this study as high mean and SD measures) in the first six months of life, in terms of f_0 variability (SD/mean), modal voice demonstrated the highest degree of control. This control occurs within the first month of life and remains for the entirety of the study. Even whilst significant changes are occurring in the anatomic-physiological structure of the infants' vocal tract and respiratory system, an infant is able to regulate the degree of f_0 variability so as to best mimic the dominant surrounding voice quality. This suggests increasing control of the larynx and vocal fold responsible for voicing.

It also suggests that previous studies' comparability issues due to methodological differences continue to need to be addressed. By focusing on just one aspect of one of the areas that [3] identified, it has been shown how voice quality contributes to the f_0 characteristics of infant vocalisations.

Table 4: *Monthly proportion of vocalisations produced with each auditory-perceptual voice quality*

Voice Quality	**Month** 1	2	3	4	5	6
Breathy	7.9%	10.5%	15.9%	10.2%	6.6%	10.2%
Creaky	21.3%	13.2%	11.8%	12.8%	14.2%	14.0%
Harsh	23.3%	15.8%	17.5%	24.7%	21.4%	24.9%
Loft	6.0%	6.3%	3.9%	3.7%	7.3%	4.6%
Modal	6.2%	9.4%	17.4%	25.8%	24.6%	22.4%
Voiceless	29.1%	39.2%	32.2%	20.0%	22.4%	20.7%
Whisper	2.1%	2.2%	0.4%	0.9%	0.9%	0.8%
Whispery Voice	4.1%	3.4%	0.9%	1.9%	2.5%	2.4%

5. Conclusions

This paper provides additional insight into the developmental trends evident in pre-babbling infants' vocalisations. Across the course of the study, no overall decrease or increase was evident in mean f_0. In addition, the overall mean and SD were similar to those found in previous studies. By including vocalisations produced with non-modal phonation, their impact on the f_0 data was able to be discerned. Individual voice qualities were able to be distinguished from one another on the basis of f_0. Different voice qualities also displayed varying amounts of variability (SD/f_0) across the entirety of the study, with modal voice showing the least variation. Whilst the growth of the vocal tract seems to have limited amounts of impact on the mean values of f_0, it does have a role in displaying the increasing control infants have over the processes used for voicing. Infants are able to regulate the degree of f_0 variability even whilst anatomic changes are occurring. The results presented in this paper support the notion that the f_0 of infant vocalisations provide insight into how an infant learns to exercise vocal control and that voice quality is a useful category through which to investigate these developments.

6. References

[1] Delack, J. B., & Fowlow, P. J. (1978). The ontogenesis of differential vocalization: development of prosodic contrastivity during the first year of life. In The development of communication (pp. 93–110). New York: John Wiley & Sons.

[2] Whalen, D. H., Levitt, A. G., Hsiao, P. L., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French-and English-learning infants. Journal of the Acoustical Society of America, 97, 2533–2539.

[3] Iyer, S. N., & Oller, D. K. (2008). Fundamental frequency development in typically developing infants and infants with severe-to-profound hearing loss. Clinical Linguistics & Phonetics, 22(12), 917–936.

[4] Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants and parents utterances in longitudinal recordings. The Journal of the Acoustical Society of America, 119(3), 1636.

[5] Buder, E. H., Chorna, L. B., Oller, D. K., & Robinson, R. B. (2008). Vibratory Regime Classification of Infant Phonation. Journal of Voice, 22(5), 553–564.

[6] Nakatani, T., Amano, S., Irino, T., Ishizuka, K., & Kondo, T. (2008). A method for fundamental frequency estimation and voicing decision: Application to infant utterances recorded in real acoustical environments. Speech Communication, 50(3), 203–214.

[7] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59(4), 390–412.

[8] Team, R. C. (2012). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

[9] Laufer, M. Z., & Horii, Y. (1977). Fundamental frequency characteristics of infant non-distress vocalization during the first twenty-four weeks. Journal of Child Language, 4(02), 171–184.

[10] Robb, M. P., & Saxman, J. H. (1989). Vocal fundamental frequency characteristics during the first two years of life. Journal of the Acoustical Society of America, 85(4), 1708–1717.

[11] Sheppard, W. C., & Lane, H. L. (1968). Development of the Prosodic Features of Infant Vocalizing. Journal of Speech and Hearing Research, 11(1), 94.

[12] Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. Journal of the Acoustical Society of America, 72(2), 353–365.

[13] Bennett, S., & Bennett, S. (1983). A 3-Year Longitudinal Study of School-Aged Children's Fundamental Frequencies. Journal of Speech and Hearing Research, 26(1), 137.

[14] Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. Journal of Speech, Language and Hearing Research, 50(6), 1510–1545.

[15] Prescott, R. (1975). Infant cry sound; developmental features. Journal of the Acoustical Society of America, 57, 1186–1191.

[16] Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. Journal of the Acoustical Society of America, 117(1), 338–350.