



Duration of early vocalisations

Adele Gregory¹, Marija Tabain¹, Michael Robb²

¹Linguistics, School of Humanities, La Trobe University, Melbourne, Australia

²Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand

am2gregory@students.latrobe.edu.au, m.tabain@latrobe.edu.au, michael.robb@canterbury.ac.nz

Abstract

The duration of three infants' vocalisations were examined during a six-month longitudinal study. In contrast to most other infant research, this study included in its analysis all vocalisations including those deemed vegetative or those having non-modal voice quality. All three infants produced vocalisations which decreased in duration in the initial months. However between the 3rd and 5th month a significant increase ($p < 0.001$) in the duration of vocalisations was found (from mean 207ms to 431ms). When vocalisations were analysed using perceptual voice quality categories, all were found to have significant differences ($p < 0.05$) in duration relative to modal voice. Non-modal voice qualities showed initial decreases in duration before increasing in duration in later months. In contrast those vocalisations produced using modal voice showed a positive linear trend and had the greatest linear rate of change across the study. These findings highlight the importance of including a wide variety of infant vocalisations including those with non-modal voice quality in infant linguistic developmental studies.

Index Terms: duration; infant vocalisation; voice quality

1. Introduction

Early infant vocalisation is viewed as the beginning of communication and linguistic development [1]. These sounds undergo numerous changes throughout the first year of life and as such various aspects of them have been subject to investigation. Studies have examined the development of a wide range of infant vocalisations such as cries [2], laughter [3] and non-distress vocalisations [4]. Research has tried to describe universal stages of speech development by documenting commonalities in early vocalisations related in part to changing anatomy, respiration and neuro-motor control [5, 6, 7]. Previous investigators who have studied non-distress and non-vegetative vocalisations have suggested that developmental changes between the 2nd and 4th month of life might impact the duration of vocalisations [8]. These developmental changes include increased neuromotor control, descent of the larynx, change in rib cage position relative to the spine and the cessation of obligate nasal breathing. The present study was designed to investigate developmental related changes in the duration of infant vocalisations during the first six months. It hypothesised that the duration of infant vocalisations would increase during between the 3rd and 5th month of life due to the developmental changes mentioned above. Previous studies have used a subset of infant vocalisations in their data [4, 9]. They have discarded those sounds deemed vegetative (such as wheezes, sneezes, coughs, hiccups and clicks) or those with non-modal voice quality comprising of unstable formant structures, highly ranging fundamental frequency or intermittent voicing. However these categories, especially non-modal voice quality, occupy a large pro-

portion of the sounds an infant produces. In an effort to more fully understand the developmental changes in early infant vocalisations, the present study included these sounds. Therefore this study aims to examine changes in duration of all infant vocalisations. Furthermore, it will also examine whether these previously excluded categories (those vocalisations with non-modal phonation), display similar duration properties and developmental trends.

2. Method

2.1. Participants

Three infants participated in a six-month longitudinal study on the laryngeal aspects of infant language acquisition. All infants were full-term with no major birth complications. They were all Caucasian (2 female, 1 male) with English speaking parents. There have been no apparent anatomical or physiological disabilities and all have normal developmental histories including normal hearing.

2.2. Procedures

Infants were videotaped in their own home beginning when they were 2-3 weeks of age. A Sony DCR-TRV16E digital video recorder with integrated microphone was used to film the infants interacting with their caregivers, or engaged in solitary play. The infants were 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 infants or be in the way of their caregivers, whilst ensuring proper capture of the sounds produced by the infants. Care was taken to minimise background noise and to provide consistency in the recording environment. Given the young age of participants (up to 26 weeks) no elicitation of segments was attempted; instead all sounds produced by the infants during the recording session were coded and labeled.

During the six-month longitudinal study, the infants were recorded on average for one hour per week. Due to the nature of the participants involved, a weekly was occasionally cut short owing to a number of reasons such as sickness or tiredness of either the infants or the caregivers. This did not affect the number of tokens available for the study as some weeks the infants were more vocal than others regardless of the length of time recorded. Once collected, the raw video was 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 labelling was conducted by the first author. Since the recordings were performed in a natural context (the infants homes) some sounds needed to be discarded due to background noise, technical diffi-

culties 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.

2.3. Coding

A vocalisation was identified as a discrete sound occurring within one respiratory cycle. Each vocalisation was labeled for voice quality (phonatory categories breathy, modal, loft, creaky, harsh, whisper, whispery voice and voiceless). Vegetative vocalisations were also assigned a voice quality based on their perceptual and acoustic characteristics and were predominantly coded as harsh, loft or voiceless. An acoustic analysis system (PRAAT version 5.1.41) was then used to determine the duration of each utterance. Inter-rater reliability for this labelling was calculated as a Kappa of 0.8 and a percentage of agreement of 83.7%. A total of 5,202 infant vocalisations were used in this study. In order to achieve more reliable estimates and to be able to compare findings with previous papers, the data were collapsed into six month-long periods.

3. Results

Results are shown in Tables 1-3 for each individual child. Segment durations in spontaneous adult speech are not normally distributed, but follow a lognormal distribution [10]. Although the infant participants were not producing 'speech' the duration of their vocalisations and individual segments were also not normally distributed (Figure 1). A log transformation of infant duration data appears more normally distributed (Figure 2). Therefore, to facilitate reliable statistical analysis, all computations were in terms of log duration. Due to this transformation the mean and standard deviation are reported as geometric mean and coefficient of variance.

Table 1: Geometric mean and Coefficient of Variance child AB

Month	Utterances	Mean	Coefficient of Variance
1	122	668 ms	201 (95% CI : 222 ms to 2008 ms)
2	247	216 ms	292 (95% CI : 55 ms to 845 ms)
3	705	174 ms	197 (95% CI : 59 ms to 517 ms)
4	661	235 ms	176 (95% CI : 85 ms to 649 ms)
5	558	357 ms	148 (95% CI : 144 ms to 885 ms)
6	538	365 ms	190 (95% CI : 126 ms to 1059 ms)

Table 2: Geometric mean and Coefficient of Variance child AP

Month	Utterances	Mean	Coefficient of Variance
1	218	205 ms	266 (95% CI : 56 ms to 750 ms)
2	269	159 ms	285 (95% CI : 41 ms to 614 ms)
3	103	129 ms	265 (95% CI : 35 ms to 472 ms)
4	173	400 ms	213 (95% CI : 128 ms to 1254 ms)
5	207	610 ms	183 (95% CI : 216 ms to 1727 ms)
6	149	576 ms	157 (95% CI : 224 ms to 1478 ms)

Table 3: Geometric mean and Coefficient of Variance child HL

Month	Utterances	Mean	Coefficient of Variance
1	161	493 ms	200 (95% CI : 164 ms to 1481 ms)
2	250	366 ms	167 (95% CI : 137 ms to 977 ms)
3	274	406 ms	160 (95% CI : 156 ms to 1054 ms)
4	238	467 ms	110 (95% CI : 222 ms to 983 ms)
5	115	575 ms	117 (95% CI : 265 ms to 1246 ms)
6	214	545 ms	99 (95% CI : 274 ms to 1084 ms)

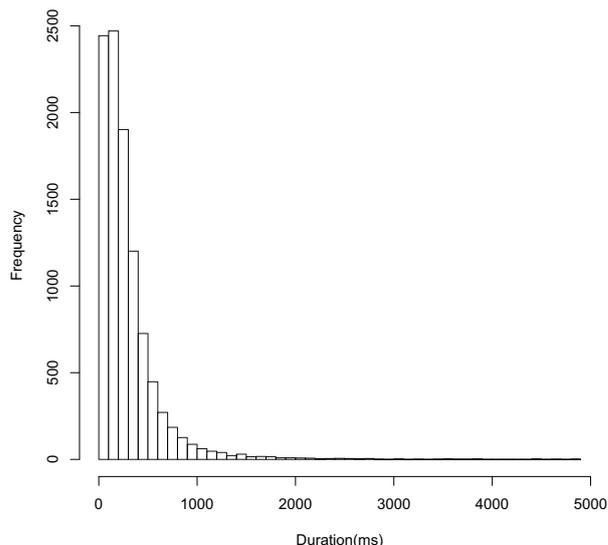


Figure 1: Histogram of durations from all infant vocalisations

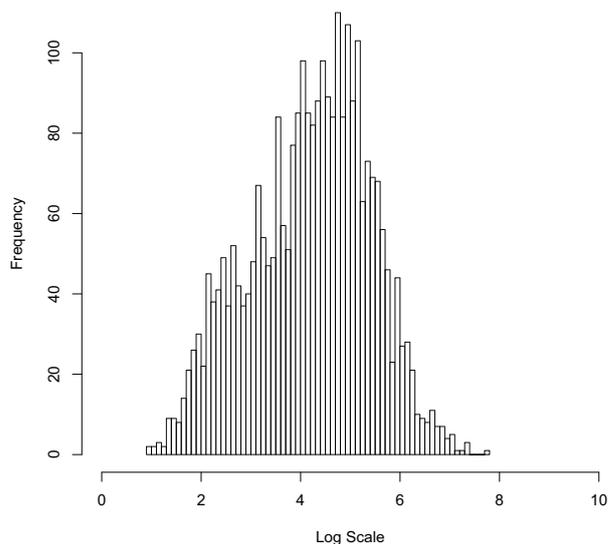


Figure 2: Histogram showing Log transformation of durations

Figure 3 presents the geometric mean against month for each child. Although the values vary for each child, what is evident is that the duration of their vocalisations actually decreases over the first two-three months before then increasing again. For child AB the first month had the longest mean duration. In HL and AP the mean duration reached its peak at the 5th month and then declined from the 6th month. Non-linear developmental changes involving both progression and regression are thus apparent in the individual data of each child. Quartic regression provides a good fit for each child ($R^2(AB)=0.995$, $R^2(AP)=0.988$, $R^2(HL)=0.971$ respectively). Previous investigators who have studied non-distress and non-vegetative vocalisations have suggested that between the 2nd and 4th month developmental changes might impact the duration of vocalisations [8]. This study hypothesised an increase in duration between the 3rd and 5th months.

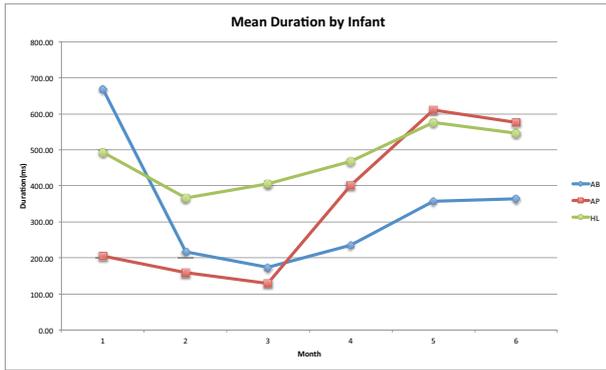


Figure 3: *Developmental changes in infant vocalisations: mean duration*

A Welch Two sample t-test was conducted to compare the combined durations for all three children in the 3rd and 5th months. There was a significant difference in duration between the 3rd (mean = 207 ms) and 5th (mean = 431 ms) month, $t(1957.648) = -15.3276$, $p < 0.001$. It is at this stage of life that control of sub-glottal air pressure is more achievable by infants as a mechanical consequence of rib restructuring [11]. Overall the durations noted in in this study are shorter than those reported elsewhere. Beebe, Jaffe, Feldstein, Mays and Alson (1985) had a mean duration (587 ms) at 3-4 months. Oller (1980) found an increase of duration from about 0.7s at three months to about 1.4s at five months. In this study even the child with the longest durations (HL) had shorter durations than those reported elsewhere. This was expected due to this study's methodology of including all sounds produced by the infants. This incorporates voiceless sounds previously disregarded, especially clicks (predominantly voiceless in quality), which have very short durations. This likely depressed the mean duration in each month.

3.1. Duration and Voice Quality

To gain a fuller understanding of the way these different voice quality parameters are interacting with the duration, the geometric mean duration was determined for each perceptual voice quality. Because a vocalisation could have more than one voice quality, the duration was calculated for each voice quality or phone alternation.

When plotted clear differences were evident between varying voice qualities. The data were analysed using linear mixed effects models. As the model is produced using linear mixed

effects it is limited to linear results, however the possibility that many voice qualities follow a polynomial regression due to their initial decrease in duration is noted. Subjects were included as random effects. The log of duration and the interaction of voice quality either side of a week 14 split were included as fixed effects. Normality and homogeneity were checked by visual inspection of residuals against fitted values and the model performed significantly better than the null hypothesis.

The modal voice quality was used as a baseline for the model. The duration of vocalisation increased linearly from a mean duration of 203 ms to 314 ms either side of the week 14 split with statistical confidence ($p < 0.001$). Relative to the modal quality, all voice qualities had significant difference in duration; creaky (144 ms greater, $p < 0.001$), whispery-voice (48 ms greater, $p < 0.023$), harsh (42 ms greater, $p < 0.001$), breathy (19 ms lesser, $p < 0.036$), loft (31 ms lesser, $p < 0.013$), whisper (49 ms lesser, $p < 0.013$), voiceless (144 ms lesser, $p < 0.001$). In addition, the modal quality showed the greatest linear rate of change (approx. 4.3 ms / week); the qualities with significant difference in slope were whisper (approx. 0.19 ms / week), creaky (approx. 0.48 ms / week), whispery-voice (approx. 1.4 ms / week), breathy (approx. 1.9 ms / week) and harsh (approx. 2.8 ms / week). The durations of these significant voice qualities are in Figure 4 with the proportion of utterance by voice quality category explained in Table 4.

4. Discussion

This study examined developmental changes in infants vocalisations during the first six months of life. Results showed that there were significant differences in the mean duration depending on the month produced. Of particular note was the increase in duration between the 3rd and 5th month. This dramatic increase coincides with significant changes in the anatomical-physiological structure of the infants vocal tract and respiratory system. Also of interest is that although durations increased for all infants across the six months, they did initially regress, declining during the 2nd and 3rd months. The reason for this initial regression needs to be more fully investigated but may be related to what vocalisation type (vegetative or non-vegetative) or what voice quality (modal or non-modal) was being produced in these early months. A polynomial (quartic) developmental trajectory provided a close correlation. A longer study would be needed to determine whether this trend seen here in the first 6 months continues into babbling and first words. The non-linear trend is especially important when seen alongside the voice quality data. Although they too increase in duration over the course of the six months, many of the voice qualities also decrease in duration over the first few months. Previous liter-

Table 4: *Mean duration of phonatory settings within vocalisations in the first six months of life (ms).*

Month	Harsh	Creaky	Whispery Voice	Whisper	Modal	Breathy	Loft	Voiceless
1	259	409	267	177	149	216	178	68
2	239	331	234	170	182	180	177	68
3	233	288	248	128	227	183	134	54
4	254	325	221	104	229	198	231	53
5	343	355	269	161	323	234	273	89
6	356	417	365	193	388	245	321	104
% of Utterances	20%	12%	2%	1%	21%	11%	4%	29%

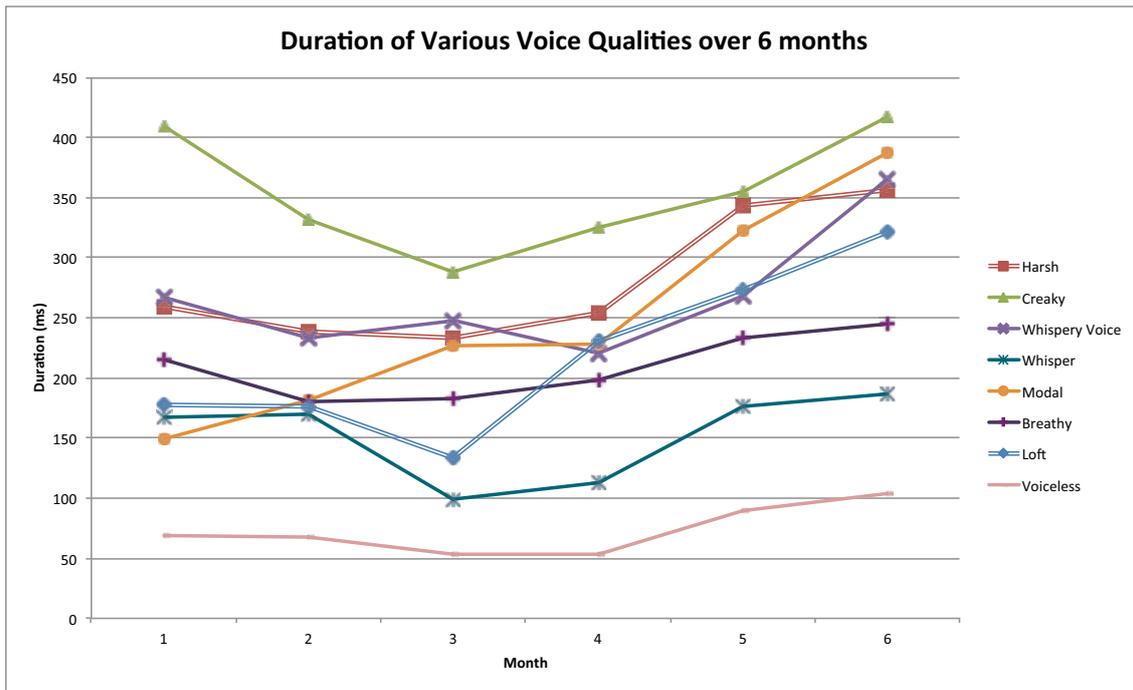


Figure 4: Duration of voice qualities over six months (ms)

ature that found positive linear increase in duration over time without this initial decrease may have done so because they did not include vocalisations with non-modal voice qualities. For when only modal tokens were examined in this study, there was a clear linear developmental trend ($R^2 = 0.932$) Figure 5. The inclusion of other voice quality modalities are therefore an important consideration when looking at duration in infant vocalisations. By excluding those that do not fit the modal parameters, it is possible to overstate the linear nature of developmental increases in duration.

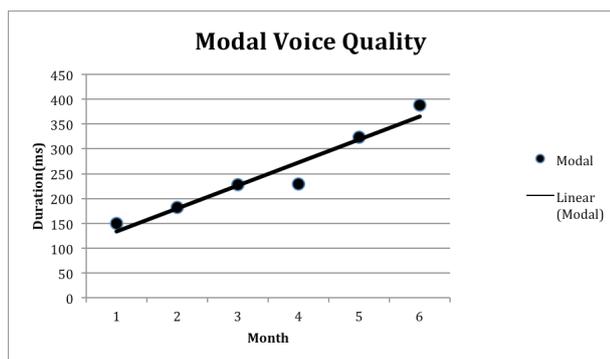


Figure 5: Duration of Modal voice quality over six months (ms)

The ability to sustain vocalisations for longer durations demonstrates an increase in the control of sub-glottal pressure. This is especially interesting in the light of the use of different voice qualities throughout this time period. In adults the use of all these voice qualities unintentionally are evidence for a disordered voice. Robb (1988) found that three parameters that characterised disorder in adults were common and typical in infants. Benner (2009) found that non-modal voice qualities along with

other laryngeal sounds are prevalent in infant vocalisations. As sub-glottal control increases, helped in part by rib restructuring, modal vocalisations increase in length. Not only are infants able to regulate the sub-glottal air pressure required for modal vocalisation but they are also able to approximate their vocal folds for regular cycles and complete closure. In contrast, the length of breathy and creaky vocalisations do not significantly increase in length over the first six months. For each of these voice qualities, durations vary but are not really affected by anatomical changes. The significant differences in duration between some voice quality categories could be understood as a function of their production. Creaky voice, as produced with tightly approximated vocal folds and irregular cycles, are physiologically easier to produce for a longer duration than breathy voice which is produced with vocal folds that do not completely close and require airflow throughout the entire cycle.

5. Conclusions

This study has examined the duration of three infants vocalisations during a longitudinal study. As a result it has shown that durations increase across this time period with significant expansion from the 3rd to 5th month. However the overall trend of the data follows a quartic regression, which takes into account the initial decrease in duration seen in all three infants. When durations were analysed using perceptual voice quality categories, it was possible to see that modal vocalisations presented a strong linear increase. Other voice qualities again showed initial decreases in duration. A number of voice quality categories showed to have significant differences in duration probably due to the intrinsic difference in their physiological production. By including all of these voice quality modalities, this study has shown the importance of including all productions that infants make without biasing or limiting the data to those tokens regarded to be speech-like.

6. References

- [1] Stark, R. E., "Features of infant sounds: the emergence of cooing," *Journal of Child Language*, vol. 5, pp. 379—390, 1978.
- [2] Wermke, K. and Robb, M., "Fundamental frequency of neonatal crying: Does body size matter?" *Journal of Voice*, vol. 24, no. 4, pp. 388–394, 2010.
- [3] Nwokah, E., Hsu, H., Dobrowolska, O., and Fogel, A., "The development of laughter in mother – infant communication: timing parameters and temporal sequences," *Infant Behavior and Development*, vol. 17, pp. 23—35, 1994.
- [4] Hsu, H.-C., Fogel, A., and Cooper, R. B., "Infant vocal development during the first 6 months: Speech quality anmelodic complexity," *Infant Child Development*, vol. 9, pp. 1–16, 2000.
- [5] Kent, R. D., "Sensorimotor aspects of speech development," in *Development of perception*, Aslin, R. N., Alberts, J. R., and Peterson, M. R., Eds. New York: Academic Press, 1981, vol. 2, pp. 161–189.
- [6] ———, "Psychobiology of speech development: Coemergence of language and a movement system," *American Journal of Physiology*, vol. 246, pp. 888–894, 1984.
- [7] Stark, R., "Stages of speech development in the first year of life," in *Child Phonology*, Yeni-Komshian, G. H., Kavanagh, J. F., and Ferguson, C. A., Eds. New York: Academic Press, 1980, vol. 1: Production, pp. 73—92.
- [8] Bloom, K., "Duration of early vocal sounds," *Infant Behavior and Development*, vol. 12, pp. 245—250, 1989.
- [9] Fagan, M. K., "Mean length of utterance beforewords and grammar longitudinal trends and developmental implications of infant vocalizations," *Journal of Child Language*, vol. 36, pp. 495–527, 2009.
- [10] Rosen, K. M., "Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison," *Journal of Phonetics*, vol. 33, no. 4, pp. 411–426, Oct. 2005. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0095447005000124>
- [11] Lieberman, P., "The physiology of cry and speech," in *Infant Crying*, Lester, B. and Boukydis, C., Eds. New York: Plenum Press, 1985, pp. 29–57.
- [12] Robb, M. and Saxman, J., "Acoustic observations in young children's non-cry vocalisations," *Journal of the Acoustical Society of America*, vol. 83, no. 5, pp. 1876–1882, 1988.
- [13] Benner, A., "Production and perception of laryngeal constriction in the early vocalizations of bai and english infants," Ph.D. dissertation, University of Victoria, <http://web.uvic.ca/ling/students/graduate/theses-dissertations.htm>, 2009.