Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 25 (2003), 197–206.

PROGRESSION IN VOWEL PRODUCTION Comparing Deaf and Hearing Children[1]

Jeannette M. van der Stelt, Ton G. Wempe & Louis C.W. Pols

Abstract

An interesting but so far neglected topic in the development of infant sound production is the hypothesized progression toward adult vowel quality. Likely, this process is quite different for normally hearing babies and for deaf babies. A band filtering analysis method is used to measure the spectral envelope in these high-pitched infants' sounds automatically. The audio material of 5 hearing and 5 deaf babies is collected monthly between the 5th and 17th month, with an additional recording at 24 months.
From each recording 50 randomly chosen utterances are digitized. In a PRAAT script criteria are set with regard to sound quality and F_0. The utterances are then analyzed resulting in spectral envelopes wherein the dependence on F_0 is minimized. Via a Principal Component Analysis a 2 - dimensional reference vowel space of all the hearing children at 24 months is determined in which individual data are projected. Preliminary results indicate that differences are found when comparing the 5th and the 24th months of hearing and deaf children.

1 Introduction

Since the 1980s infant vocalizations are considered to be precursors of later speech (Lindblom & Zetterström, 1986), but the developmental process is not yet unraveled. This is at least partly due to the complexity of infant vocal behavior. The usual choice in research is to label the vocalizations perceptively, but already for the 'simple' utterances of the first 6 months, coding shows a low inter-listener-reliability (Nathani & Oller, 2001). Of course, the reliability ratings are related to the simplicity or the complexity of the coding system (Koopmans-van Beinum, Clement & Van den Dikkenberg-Pot, 2001). Serkane, Schwartz, Boë, Davis & Matyear (2002) suggest two mechanisms for speech development: "exploration of the vocal tract sensori-motor abilities, and the imitation (overt stimulation) of caretakers' language sounds." (p. 45). In earlier work (Koopmans-van Beinum & Van der Stelt, 1979; 1986) we found a predictable, possibly universal development with regard to coordination of respiration, phonation and articulatory movements in infancy. In that sensori-motor system the focus was on movements and not so much on sound description itself. This articulatory definition of (canonical) babbling became an important tool in diagnosis of deafness: deaf infants have a delayed babbling onset (Koopmans-van Beinum et al., 2001; Oller & Eilers, 1988). This finding is congruent to the second mechanism proposed by Serkane et al. (2002). Deaf infants lack the possibility to imitate the adult acoustic examples.

[1] This is an extended version of the paper presented at the 15th ICPhS in Barcelona, Spain (Van der Stelt et al, 2003).

In the present study we investigate unlabeled deaf and hearing infant vocalizations acoustically in an automatic way. We assume that an anatomical change in the oral cavity is related to the use of that space as a resonator in sound production. At the same time, the use of the oral space very likely is driven perceptually as well. Thus, deaf and hearing babies will differ, possibly right from an early onset and certainly when they grow older. Yet, we have to reckon with enormous variability in developmental processes per child and over the deaf and normally hearing children. This phenomenon clearly complicates diagnosis.

2 Phonetic data

The participants in our study are 5 healthy born hearing boys (NH) matched to 5 boys born deaf (HI). Matching criteria were e.g. sex, birth order, living in and originating from the same geographical region (e.g. Van den Dikkenberg-Pot & Van der Stelt, 2001). Specific information on the hearing status of the HI children (and communication method) is given in Table 1.

The parents made monthly audio recordings of at least 30 minutes mother-infant interaction in naturalistic home situations, using the audio tracks of a video recorder with CD-quality sound recording. For this study we selected the recordings from the 5^{th} to the 17^{th} month, with an additional recording when the children were 2 years of age. In each recording 10 minutes were selected of ongoing vocal mother-infant interaction.

Within that selection 50 utterances are randomly selected with speech motor milestones in mind (Koopmans-van Beinum & Van der Stelt, 1986). So, the vocalizations were produced mostly during egressive respiration, with (interrupted) phonation, with or without one or more articulatory movements.

Some overall statistical data are available for the first and the second year of life, respectively (Clement & Koopmans-van Beinum, 1999; Van den Dikkenberg-Pot & Koopmans-van Beinum, 1997; Koopmans-van Beinum et al., 2001). In the first year of life the mean number of utterances differed significantly for the NH and the HI infants; the latter produced more vocalizations than the normally hearing infants. With regard to the mean utterance duration in the first year, a significant difference between the two groups is found as well. This effect is mainly due to the long utterance durations in month 3.5 for the NH infants. For the recordings in the second year of life the mean number of spoken utterances per recording is not significantly different for the two groups of children. The mean duration however is significantly longer for the HI children than for the NH children.

The database used in this study thus consists of 10 children x 14 monthly recordings x 50 utterances = 7000 utterances.

Table 1: Loss characteristics of the deaf children (HI). Language: O = Oral method, TC = Total Communication, DSL = Dutch Sign Language.

Subject /Lang.	Hearing loss best ear (dB)	Loss with hearing aid (dB)	Diagnosis at age (mnth)	Hearing aid from age (mnth)
HI1/O	97	55	1.5	2.0
HI2/TC	93	55	3.0	3.5
HI3/O,TC	110	65	4.0	4.5
HI4/DSL	>120	not tested	0.5	no aid
HI5/DSL	120	not tested	3.0	6.5

3 Acoustic Analysis

Estimating formants objectively in high-pitched sounds is notoriously difficult (e.g. Wempe, 2001; Vallabha & Tuller, 2002). The vocal tract functions as a filter for the source sound. High pitch causes undersampling of the spectral envelope, whereas source and filter functions, strictly spoken, cannot be separated. Many procedures have been developed, of which the LPC inverse filtering method is accepted as a reasonable compromise. Researchers usually select the vowels or syllable-like sounds perceptively and label them phonologically, which permits them to decide about the order of LPC analysis and the most probable formant positions. This procedure however biases the formant detection subjectively, and wrongly chosen parameters could easily produce very misleading results (e.g. Wempe, 2001; Vallabha & Tuller, 2002).

In order to be able to compare vocalizations of 5-month-old infants (born deaf as well as normally hearing) all the way up to the age of two, Wempe (2001) has developed a pitch-related band filter analysis via scripts in the PRAAT-program (e.g. Wempe & Boersma, 2003).

First, we analyzed the distribution of the pitch (F_0) in all voiced utterances. The inter-child variability in the utterances over the recordings was considerable. In all children the range is very large (about 100-800 Hz). Per child and per recording the mean F_0 varied as well, but remained mainly below 400 Hz. Based on this information we decided about the bandwidth of the filter (see below). In Figure 1 the overall mean and s.d. are given for the mean F_0 per recording of the 5 HI and the 5 NH children.

Secondly, a script is written ultimately aiming at an estimation of a spectral representation of an utterance. The first module in that script divides each vocalization in 10 parts of equal duration, which permits a distribution of measurement positions over the entire utterance, thus covering a possible articulatory change within that vocalization. Bandpass Filter Analysis, described below, is done in all 10 parts of each sound, theoretically resulting in maximally 500 measurements per month if all parts are voiced and meet the other criteria as well. The second module is concerned with sound quality: clipped and low-intensity parts, possibly representing consonantal parts of the vocalizations, are avoided (max. <-1dB and min. >-10 dB relative to the absolute peak level within the utterance).

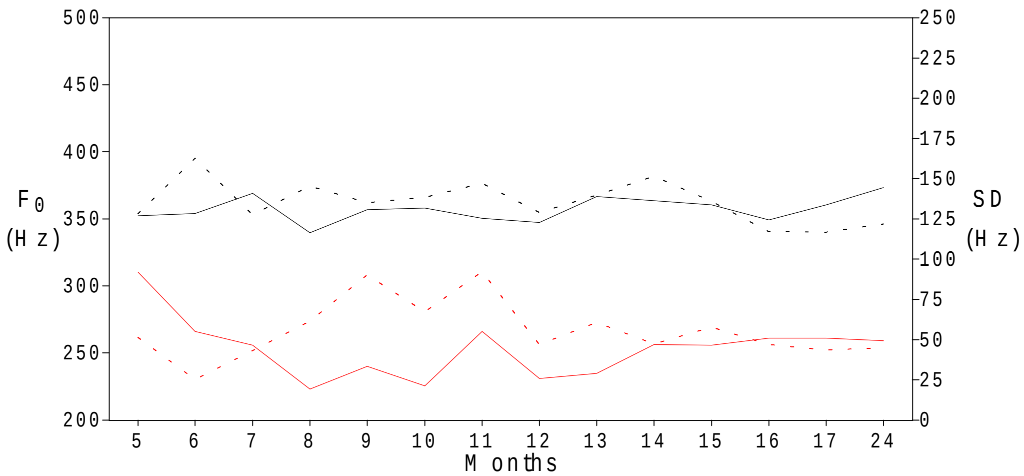

Figure 1: Distribution of F_0 (upper set) and s.d. (lower set) in utterances per recording, continuous lines are used for the NH children, dotted for the HI. The left Y-axis gives frequencies for mean pitch, the right Y-axis for the s.d. values.

The Bandpass Filter Analysis (BFA) is applied to those (maximally 10) parts of the utterance where F_0 is below 425 Hz for at least 3 contiguous pitch frames. In the third module of the script one period is selected in each valid part of the sound. This period is recycled up to a duration of 0.1 second. This artificial sound is multiplied with a Gauss-window (10 Hz width) and pre-emphasis is used. Then a swept Gaussian bandpass filter analysis (step = effective BW/5, with an effective BW=1.1 x 425 Hz) is done. The choice of this filter bandwidth is a compromise between frequency resolution and F_0 - ripple. Furthermore, an RMS level normalization to 0.3 Pa is applied to each measurement, which avoids energy variance caused by, sometimes big, differences in recording levels.

An intensity contour can be plotted representing a bandfilter spectrum covering a range from 0 to 7 kHz. Linear rather than logarithmic spectra of each measurement are used for further analyses. The bin widths are set to 175 Hz resulting in 40 values per measurement. Ideally, per utterance 400 values are calculated (10 parts x 40 bins) and collected. Per child and per month a matrix is produced with the bin energies. Matrices thus represent the intensity in each of the 40 filters over the total number of measurements from a monthly recording, ideally resulting in 10 parts * 50 utterances = 500 values per frequency bin. The actual number of measurements per monthly recording is calculated and given in the matrix as well. In practice, the selection criteria (concerning F_0 and intensity) limit this number to 38% on the average so that about 200 values per bin per monthly recording remain for further processing. The total number of values thus amounts to about 38% of 500 values * 40 bins * 14 months * 10 children = 1.06 million. Each matrix represents a bandfilter spectrum covering a range from 0 - 7 kHz. In the spectra, maxima could be detected automatically, but this will not give consistent results either. That is why we choose the whole-spectrum approach and data reduction via PCA. Each bandfilter matrix then can be regarded as a point in a 40-dimensional space. By using the first two principal components the data are reduced and can easily be displayed.

The above scripts, subdivided in several modules, are applied to the digitized utterances of all 10 children at all 14 time periods.

4 Results

In order to compare deaf and hearing children in their development towards an adult vowel space, we constructed a 'reference plane'. The 40-dimensional spectra of all utterances at 24 months of the 5 hearing children are merged into one table and reduced by means of a Principal Component Analysis (PCA). In Figure 2 the overall distribution of variance is given. The weights of the first and second eigenvector are given as well.

Table 2: Percentage of variance explained by each of the first 5 vectors and cumulative as well in a PCA over the variance-covariance matrix of all analyzable utterances of the 5 two-year-old NH children (see also Figure 2).

		Cumulative
eigenvector 1	32.5 %	32.5%
eigenvector 2	22.0 %	54.5%
eigenvector 3	10.2 %	64.7%
eigenvector 4	9.3 %	79.0%
eigenvector 5	5.6 %	79.6%

As can be seen from Table 2, the first and second PC together explain 54.5 % of the variance. All 855 measurements of the 5 NH children are projected in a plane constructed by means of these first two PCs (see Figure 3). The first and the second eigenvectors are used to draw a plane in which all bin spectra are displayed as points. This PC1-PC2 plane is often taken as an analogue to the F_1-F_2 plane.

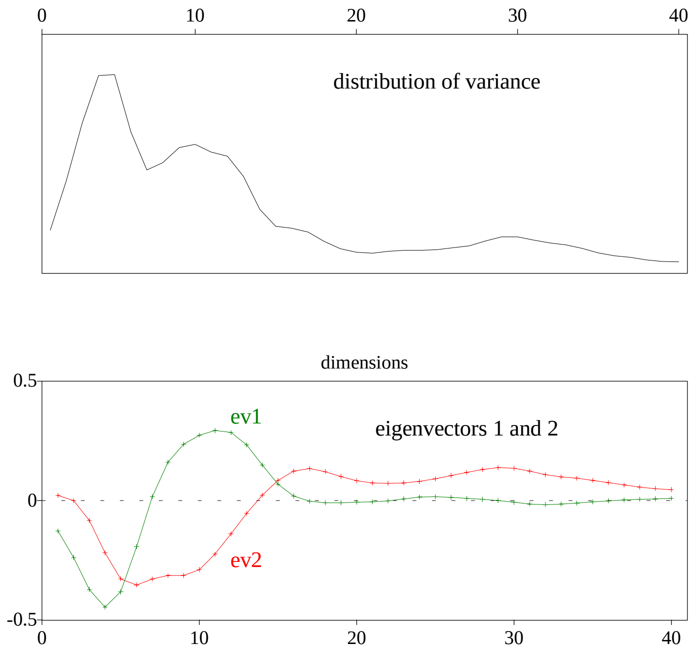

Figure 2: Graphs of the distribution of variance and the contribution of the first 2 eigenvectors after PCA of all filter values for the 5 NH children in the 24th month.

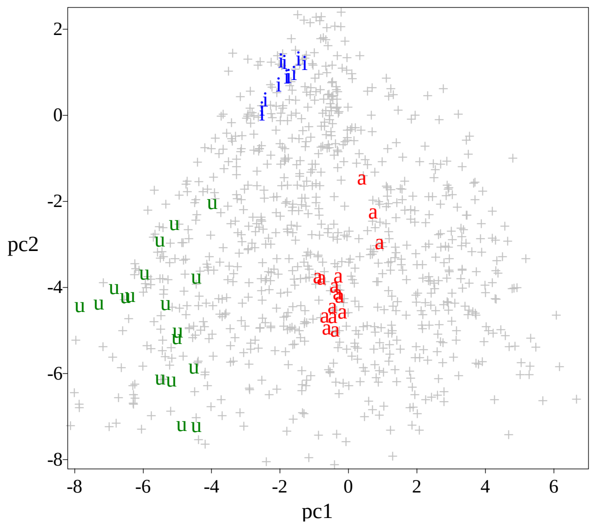

Figure 3: Normal reference vowel space for all analyzable utterances of the 5 normally hearing children (gray +). Labeled vowels from an adult female speaker are projected in this plane as well (/i/, /u/, and /a/).

In the above-described manner the distribution of points in Figure 3 can be considered as the vowel space of normally developing two-year-olds. However, the utterances were unlabeled and therefore we added some adult labeled vowels for reference as well. These labeled vowels originate from isolated words of a female speaker and were analyzed similarly to the procedure used for the utterances of the 10 children. As can be seen, the vowels are almost distributed like in an adult vowel formant space (although in a somewhat different direction), and they all fall within the children's space. This is congruent with growth models, and can be explained by considering the changing vocal tract qua size and shape in young children.

We have further checked this assumption by manipulating the labeled adult vowels (see Figure 4) in such a way that the sounds become more childlike. The F_0 of the female speaker was about 224 Hz. In order to meet the criterium that F_0 must be below 425 Hz we could override the sample frequency of the female speaker with a factor 1.8. The sample frequency changed from 22050 to 39690 Hz, thus changing the formants to the childlike domain. Comparing Figures 3 and 4 we conclude that the vowels /i/ and /u/ remain in approximatedly the same region, whereas the vowel /a/ moves more upwards.

The reference vowel space we constructed with all analyzable utterances of the 5 normally hearing children when 24 months old clearly resembles the F_1-F_2 plane. The bigger change in the location for the /a/ vowels may be explained by findings of Perkell and collegueas (Perkell & Cohen, 1989; Perkell & Nelson, 1985). They found that the Stevens-House speech synthesis model (1955) is also plausible for human speech motor control: variance in repeated productions of the same vowel is minimized near a vocal tract constriction. The vowel /a/ can be characterised as an open vowel. Hasagawa-Johnson, Pizza, Alwan, Setsu Cha & Haker (2003) found that in adults, intertalker variance is less for palatal and velar vowel productions than for uvular and pharyngeal vowels. Assuming that place of articulation restricts vowel production in an individual speaker as well, this possibly can explain the larger spreading for the /u/ sounds than for the /i/ sounds in both Figure 3 and Figure 4.

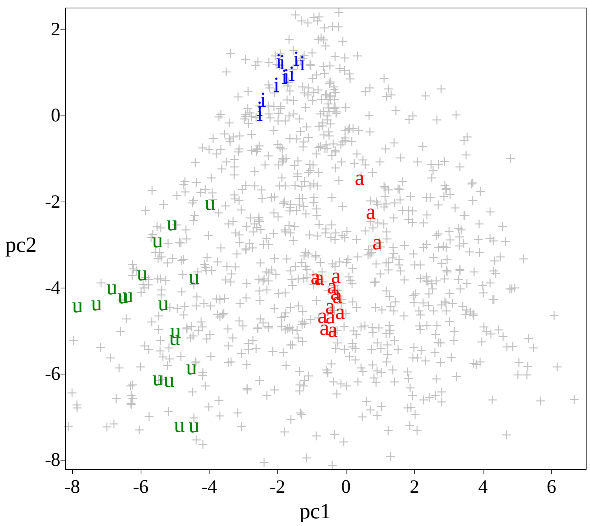

Figure 4: Normal reference vowel space for all analyzable utterances of the 5 normally hearing children (gray +). Labeled vowels from an adult female speaker are projected in this plane as well (/i/, /u/, and /a/). The sample frequency is multiplied with a factor 1.8 which resulted in an F_0.still below 425 Hz while the formants become more childlike.

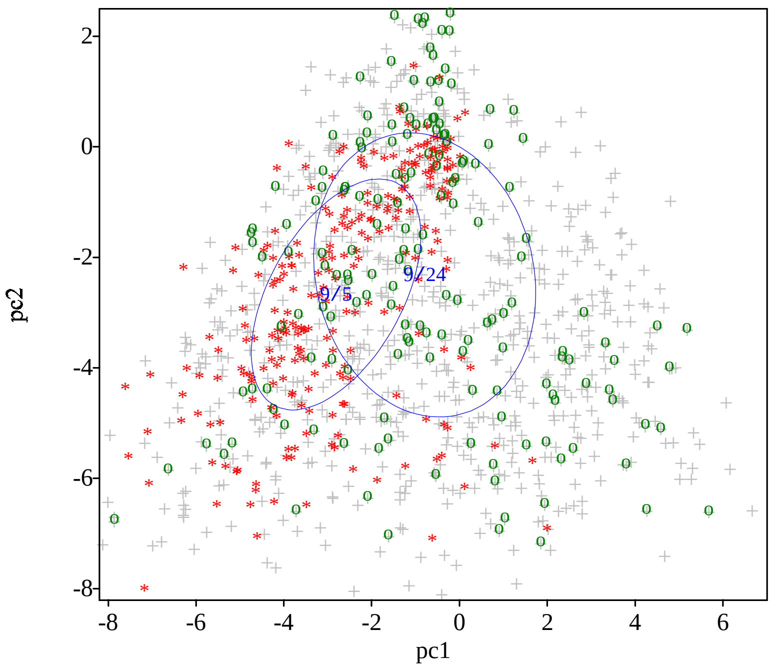

Figure 5: Reference plane with the utterances of a NH child (subject 7) when 5 (*) and when 24 months old (o) projected into that plane. The ellipses represent 1 s.d. of the variance of the 40 bins for pc 1 and pc 2 for the respective months

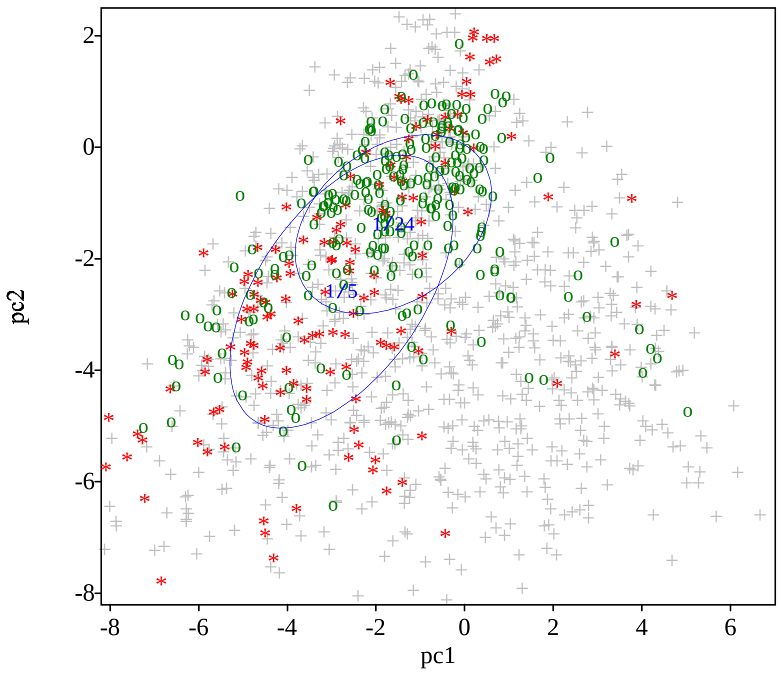

Figure 6: Reference plane with the utterances of a HI child (subject 5) when 5 (*) and when 24 months old (o) projected into that plane. The ellipses represent 1 s.d. of the variance of the 40 bins for pc 1 and pc 2 for the respective months

The data per child and per month are calculated. For this presentation, we display only the months 5 and 24 from a NH child (subject 7), and the months 5 and 24 from a HI child (subject 5) into the 24-months reference plane, see Figures 5 and 6 respectively. Of course, the data from the NH child at 24 months are part of the reference plane.
In the Figures 7 and 8 the 1 s.d. ellipses are given for all children for the 5^{th} and the 24^{th} month. Five month-old children do not differ very much, except perhaps subject 9 (NH). At 24 months of age we see a difference between NH and HI children with regard to the placement of the ellipses. For the NH children there is a considerable overlap. This overlap is comparable that of the two HI children (subjects 1 and 2), who have a hearing loss in the best ear of 97 and 93 dB respectively, and were fitted with a hearing aid relatively early (see Table 1). HI subjects 1 is educated orally and his vocal output is comparable to that of the NH children. HI subject 2 is a TC child. HI child 3 is an O/TC child and HI subjects 4 and 5 use DSL. These last two HI children do not really change with regard their vocal output between 5 and 24 months.

In Table 3 the averages and standard deviations are given per child for the 5^{th} and the 24^{th} month, for the first two eigenvectors as well as for the accumulated eigenvectors 1 and 2.

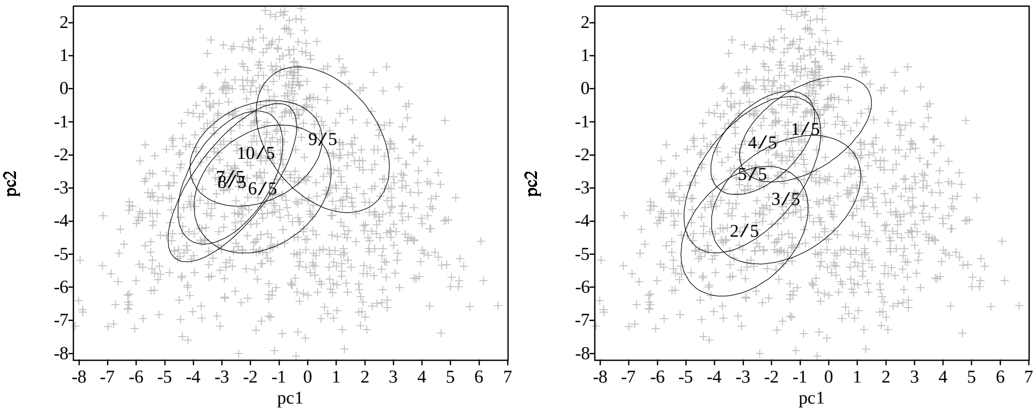

Fig. 7: Average standard deviation over the 40 bins over pc 1 and pc 2 for all children in the recording of the 5^{th} month. The HI children are represented on the right (subjects 1-5), the NH children on the left (subjects 6-10).

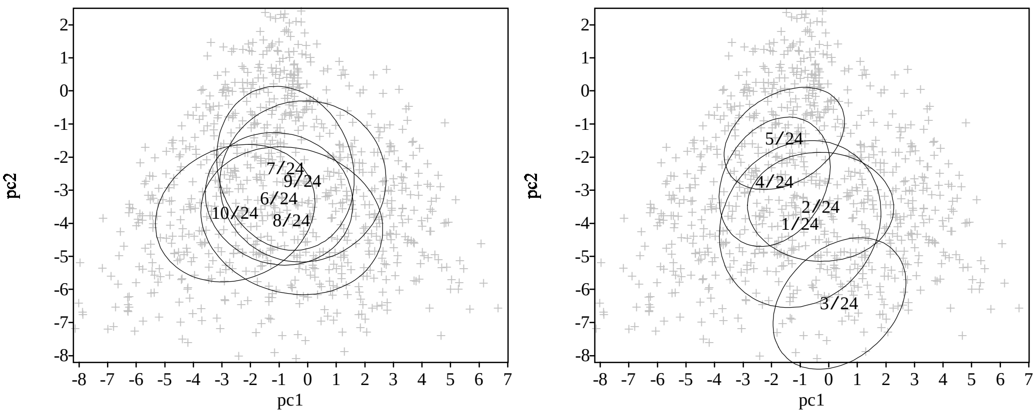

Fig. 8: Average standard deviation over the 40 bins over pc 1 and pc 2 for all children in the recording of the 24^{th} month. The HI children are represented on the right (subjects 1-5), the NH children on the left (subjects 6-10).

Table 3: Averages and standard deviations per child for the 5th and the 24th month, for the first two eigenvectors as well as for the accumulated eigenvectors 1 and 2.

month / child	5	6	7	8	9	10	11	12	13	14	15	16	17	24
NH1	3,36	2,73	3,20	3,12	2,34	2,59	1,91	2,97	2,37	2,63	2,86	2,68	2,28	2,61
NH2	3,26	3,02	2,68	2,71	2,92	3,00	2,88	3,22	3,40	3,06	2,83	2,96	2,84	3,06
NH3	2,80	2,39	2,51	2,88	2,91	2,82	2,37	2,81	2,40	2,75	2,70	2,93	2,41	3,79
NH4	2,97	2,54	2,82	3,23	3,22	3,74	2,67	3,04	3,34	3,04	2,89	3,67	3,65	3,04
NH5	2,40	2,56	2,78	2,80	2,93	3,48	2,60	2,63	2,66	3,07	2,88	2,98	2,54	2,76
HI1	3,29	2,98	3,22	3,63	3,75	2,99	3,20	3,81	4,00	4,08	3,62	3,44	4,04	3,89
HI2	3,21	2,38	2,63	2,36	2,69	2,74	2,97	2,72	2,59	2,72	3,26	2,68	2,65	3,79
HI3	3,08	3,00	2,73	2,88	2,96	2,88	2,76	3,09	2,88	2,82	2,96	3,13	3,44	3,27
HI4	2,72	3,10	3,48	2,77	3,44	3,49	3,30	3,51	2,86	3,33	3,27	2,84	2,69	3,46
HI5	2,81	2,57	3,12	3,19	3,03	3,46	3,26	3,25	2,70	2,85	2,74	2,96	2,87	3,48

5 Preliminary Conclusions

In all monthly recordings of our NH and HI subjects we found a large range in F_0 (100-800 Hz). Papaeliou, Minadakis & Cavouras (2002) found different acoustic patterns in 7- to 11-month-old children: durations differed, and peak and final F_0 are higher for the expression of emotion than for utterances with a communicative function. The s.d. in F_0 did not differ for these types of vocalizations. In our data (Figure 1) the s.d. in that age period is much larger for the HI children than for the NH. Possibly, our HI children, then, express more emotions than the NH children. Yet, with regard to the mean distribution of F_0 no significant difference is seen between the deaf and hearing children over the monthly recordings.

The data on vowel space development (5 and 24 months) indicate that for all 5 NH children the data points expand towards the normal reference space for 24-month-old children. This is no surprise since the 24-months data:were used to construct the reference plane. For the HI children this expansion is less obvious.

In the 24th month most points are still in the same area as found for the 5th month. In both children we see a concentration along the line between the adults' /u/ and /i/ projections: the closed vowels. The open vowel area /a/ possibly permits more variance in voiced sound production.

When comparing the HI and the NH children (age 5 months) with regard to the 1 s.d. elipses, we see somewhat more differences between the HI children than between the NH children, except for subject 7 perhaps. When the children are 24 months old the overlap bewteen NH children is more consistent than for the HI children. Yet, the children with the most severe hearing loss (HI 3, 4, and 5) are more comparable to the NH children than the HI subjects 1 and 2, with relatively good prognosis (loss in dB, age of detection, and hearing aid fitted).

We, preliminary, have to conclude that children during the first two years follow quite different pathways in their vocal development.

6 Further Research

These results point to various interesting aspects for further research. We will look for differences in durations and F_0 changes in the utterances in our database, since

differences between NH and HI children were found in earlier reseach (e.g. Van den Dikkenberg-Pot & Koopmans-van Beinum, 1997). Furthermore, the development of the vowel space over the different months in individual children is possibly quite interesting as well. Maybe we have to investigate the eigenvectors per child and per recording as well in order to check the amount of explained variance.
Zajdó and Stoel-Gammon (2003) have presented data on the development of vowel production in 80 Hungarian children from two to four years of age. These vowel productions are labeled and thus interesting for comparison with our unlabeled utterances of the deaf and hearing children when 24 months old. Together we intend to disseminate new results on the Dutch and Hungarian vowel productions of young children.

References

Clement, C.J. & Koopmans-van Beinum, F.J.: Characteristics of vocalizations of deaf and hearing infants in the first year of life, *Proc. ICPhS* **99**, Vol. 3, pp. 1929-1932, 1999.

Van den Dikkenberg-Pot, I. & Koopmans-van Beinum, F.J.: Number and duration of spoken utterances of deaf and normally hearing children between 12 and 18 months, *Proc. Inst. Phon. Sc., Univ. of A'dam*, **21**, pp. 91-100, 1997.

Van den Dikkenberg-Pot, I. & Van der Stelt, J.M.: Mother-child interaction in two year old deaf and hearing children, *Proc. Inst. Phon. Sc., Univ. of A'dam*, **24**, pp. 1-13, 2001.

Hasagawa-Johnson, M., Pizza, S., Alwan, A., Setsu Cha, J. & Haker, K.: Vowel Category Dependence of the Relationship Between Palate Height, Tongue Height, and Oral Area, *JSLHR*, **46**, pp.738-753, 2003.

Koopmans-van Beinum, F.J., Clement, C.J. & Van den Dikkenberg-Pot, I.: Babbling and the lack of auditory speech perception: A matter of coordination?, *Developmental Science*, **4**, pp. 61-70, 2001

Koopmans-van Beinum, F.J. & Van der Stelt, J.M.: Early Stages in Infant Speech Development, *Proc. Inst. Phon. Sc., Univ. of A'dam*, **5**, pp. 30-43, 1979.

Koopmans-van Beinum, F.J. & Van der Stelt, J.M.: Early Stages in the Development of Speech Movements, In: B. Lindblom & R. Zetterström, *Precursors of Early Speech*, Stockton Press, New York, pp. 37-50, 1986.

Lindblom, B. & Zetterström, R. (Eds.): *Precursors of Early Speech*, Stockton Press, New York, 1986.

Nathani, S. & Oller, D.K.: Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations, *Behavior Research Methods, Instruments, & Computers* **33 (3)**, p. 321-330, 2001.

Oller, D.K. & Eilers, R.E.: The role of audition in infant babbling, *Child Development* **59**, pp. 441-449, 1988.

Papaeliou, C., Minadakis, G. & Cavouras, D.: Acoustic Patterns of Infant Vocalizations Expressing Emotions and Communicative Functions, *JSLHR* **45**, pp. 311-317, 2002.

Perkell, J.S. & Cohen, M.H.: An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/, and /u/, *J. of Phonetics*, **17**, pp.123-133, 1989.

Perkell, J.S. & Nelson, W.L.: Variability in production of the vowels /i/ and /a/, *JASA*, **77**, pp. 1889-1895, 1985.

Serkane, J., Schwartz, J.L., Boe, L.J., Davis, B. & Matyear, C.: Motor specifications of a baby robot via the analysis of infants' vocalizations, *ICSLP Denver*, pp. 45-48, 2002.

Stevens, K.N. & House, A.S.: Development of a quantitative description of vowel articulation, *JASA*, **27**, pp. 401-493, 1955.

Van der Stelt, J.M., Wempe, T.G. & Pols, L.C.W.: Progression in infants' vowel space: An analysis of deaf and hearing infants'sounds, *Proc. ICPhS**2003***, Vol. III, pp. 2225-2228, 2003.

Vallabha, G.K. & Tuller, B.: Systematic errors in the formant analysis of steady-state vowels, *Speech Communication* **38**, pp. 141-160, 2002.

Wempe, A.G.: F_0-related formant measurements, *Proc. Inst. Phon. Sc., Univ. of A'dam,* **24**, p. 167-187, 2001.

Wempe, A.G. & Boersma, P.: The interactive design of an F_0-related spectral analyser, *ICPhS**2003***, Vol. I, pp. 343-346, 2003.

Zajdó, K. & Stoel-Gammon, C.: The Acquisition of Vowels in Hungarian: Developmental Data, *Proc. ICPhS**2003***, Vol. III, pp. 2229-2232, 2003.