ACOUSTIC ANALYSES OF THE VOCALISATIONS OF A HEARING-IMPAIRED INFANT: EXPLORING METHODS

Astrid van Wieringen and Florien J. Koopmans-van Beinum

ABSTRACT

To say that acoustic measurements on high-pitched voices involve many difficulties, is asking the obvious. Nevertheless, the need of acoustic data on infant voices is growing as the interest in speech developmental processes increases. Analysis of the acoustic signal is a necessary means to determine the variability and similarities within normal speech development as well as in disorders. The availability of a number of audio-recordings of a hearing-impaired infant coincided with our need to test and select reliable analysis systems, suitable for the acoustic analysis of the sound productions of infants and young children. In this pilot study we concentrated on the analyses of formant frequencies of the vocalic parts of the infant's sound productions.
Approximately 300 non-cry vocalisations of a hearing-impaired boy at the age of 9, 12, and 15 months were analysed spectrographically, and by means of a conventional as well as a robust method of linear prediction coding. In comparing our results based on these three techniques mutually and with data from literature we have to conclude that the robust linear prediction method yields extremely high second formant frequency values, whereas the other two techniques agree within a relatively small frequency range.
As for the achieved formant frequency data of the infant we feel that at this stage it is not clear yet which aspects are the result of the analysing techniques under test and which of the vocal development of the infant. Nevertheless, a satisfying agreement is found with data from literature.

1. INTRODUCTION

There has been a considerable interest in studies on infant vocalisations, either in terms of relating the vocalisations to later linguistic behaviour or in terms of describing the physiological sounds as such, as part of a total development, especially owing to the importance of detecting communication disorders at the youngest possible age (for an overview see Gilles, Koopmans-van Beinum & van der Stelt, 1988).
Whatever the motive, the study of infant speech is always approached rather hesitantly as we lack a thorough understanding not only of the neurological stages the infant undergoes in its first year, yet also of the motor structures underlying the articulatory development. To this end, a phonetic coding system based on phonation aspects and on articulatory movements of infants during their first year of life was developed (Koopmans-van Beinum & Van der Stelt, 1979 and Van der Stelt & Koopmans-van Beinum, 1981). The analyses of phonation and articulatory movements resulted in a coding scheme, which differentiates between five clear stages in the infant speech movements before the child arrives at the one-word stage (Koopmans-van Beinum &

Van der Stelt, 1986). The disadvantage, however, of the former procedure is that it is rather subjective, dependent only on those transcribing the recordings. What one would like to see is straightforward, unambiguous acoustic data. However, the main drawback to measuring vocalic features acoustically has been the lack of knowledge with respect to which features should actually be taken into account and what technique can be applied reliably to them.
Acoustic analysis is a major problem for infant vocalisations as spectral analyses are performed on a much shorter pitch period than in adults. With increasing fundamental frequency the resolution of the formants can be poor: the larger the interval between the harmonics the more difficult to estimate formant frequencies from an imaginary spectral envelope. Most acoustic studies on infant vocalisations rely on spectrograms for measuring duration, fundamental frequency pattern or formant frequencies (e.g. Kent and Murray, 1982; Kent et al. 1987; Oller et al. 1985; Oller and Eilers, 1988; Robb and Saxman, 1985, 1988). Even though spectrograms are an excellent means of representing speech visually, it remains difficult to estimate the resonances of the vocal tract accurately (within 50 Hz) from the broad black bands.

2. DESIGN

2.1 Subject

Our first and main concern of this pilot study is to find out if and how we can obtain reliable spectral measurements from infant vocalisations. For this purpose our data were submitted to digital spectrography as well as to two linear predictive coding (LPC) techniques, one being a conventional LPC, the other a so-called robust LPC method.

The data for analysis happened to be vocalisations of a hearing-impaired baby boy. We are aware that these are not fully comparable to the data of normal hearing infants. Since our recordings were from 9 months onwards, we estimated formant frequencies at 9, 12, and 15 months, making our results more or less comparable with the data from Kent et al. (1987). Their data included the utterances of a hearing-impaired infant at 8, 12, and 15 months. Our analyses can also be compared to those of Oller and Eilers (1988) concerning hearing-impaired infants between the age of 11-14 months. Once a method is developed we not only wish to analyse normal hearing infants systematically, yet we also hope to focus on an early diagnosis of communication disorders, mainly with respect to hearing-impaired infants.

2.2 Data

Our infant, having a profound hearing loss bilaterally, was audio-recorded by his mother from nine months onwards at least every month. His hearing-impairment was discovered at birth, family and medical staf being on their guard because of an hearing-impaired elder sister. Although he received hearing aids as soon as possible after birth he could not often wear them as he frequently suffered from glue ears, for which he received medical treatment. Figure 1 is the audiogram of the hearing-impaired infant at 18 months. His hearing-loss for both ears is on average 60 dB without and on average 35 dB with hearing-aids.
Approximately fifty non-cry comfort utterances per month were selected from the audio-recordings and subsequently digitized at a sample frequency of 20 kHz (cut-off frequency filter of 9.6 kHz) on a micro VAX II mini-computer. As the audio-recordings had been made at home by the infant's mother, many utterances were

unsuitable for analyses due to environmental noises (a spoon scraping in a plate, their dog barking, etc.). Of the remaining vocalisations the vowel-like sounds were used (no laughs).

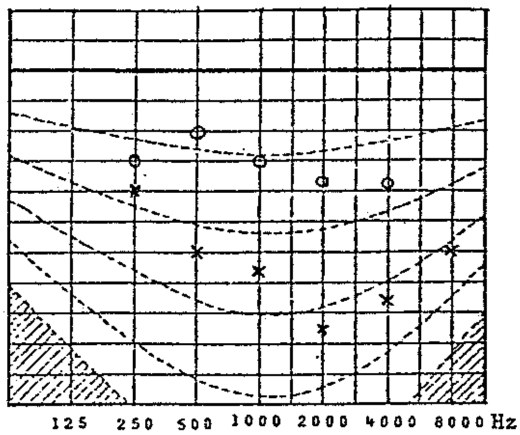
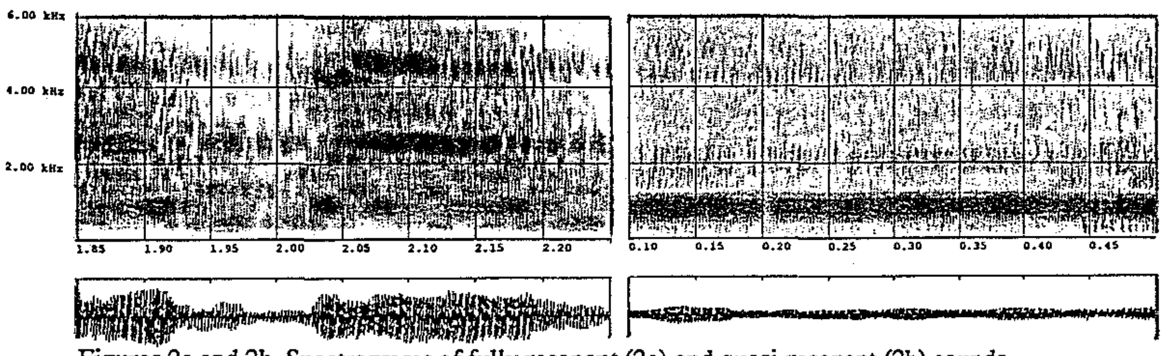


Figure 1. Audiogram of our hearing-impaired infant. Without aids his hearing-loss is on average 60 dB (X), with aids it is reduced to 35 dB (0).

The data of the 9th and 12th months differ from those of the 15th month in that the former contain no babbles at all, whereas the latter do. As would appear in the course of analyses, very infantile utterances -ones which lack any adult-like reference- are the most difficult of their kind to measure. In addition to the lack of knowledge on the estimates as a result of the analysing techniques, the structure of the utterances may differ to such an extent from that of older (and well-hearing) infants that the acoustic features of the older subjects may not be present in the data at all. As a result of the physiological maturation of the larynx, the acoustic and articulatory characteristics of early infant vocalisations change rapidly. The anatomical changes do not only include the lowering -and growth- of the larynx, but also the possibility to raise the velum and the separation of the velum and the glottis between 4-6 months of age. After this decoupling has occurred normal hearing infants will produce less nasal vowel-like utterances and more fully-resonant nuclei (Kent, 1982). In spite of a continuing physiological development, the vocalisation progress of hearing-impaired infants will lag behind here as their feedback is severely reduced.

The nuclei of early unarticulated vowel-like sounds are thought to be present in spectrograms as a "quasi-resonant" feature (Oller, 1985). As the vocal tract does not serve as a full resonator there will be a predominance of low-frequency energy only. Figure 2a and 2b illustrate a segment of a fully resonant and "quasi-resonant" sound in terms of Oller et al. (1985). We must emphasize, however, that, anatomically, our infant had passed the early vocalisation stage. As the production apparatus is just as mature as with a normal hearing infant the acoustical features observed cannot be attributed to the anatomical growth, yet only to the use of the vocal tract. It is important not to confuse the effect of hearing-impairment with that of incomplete growth.



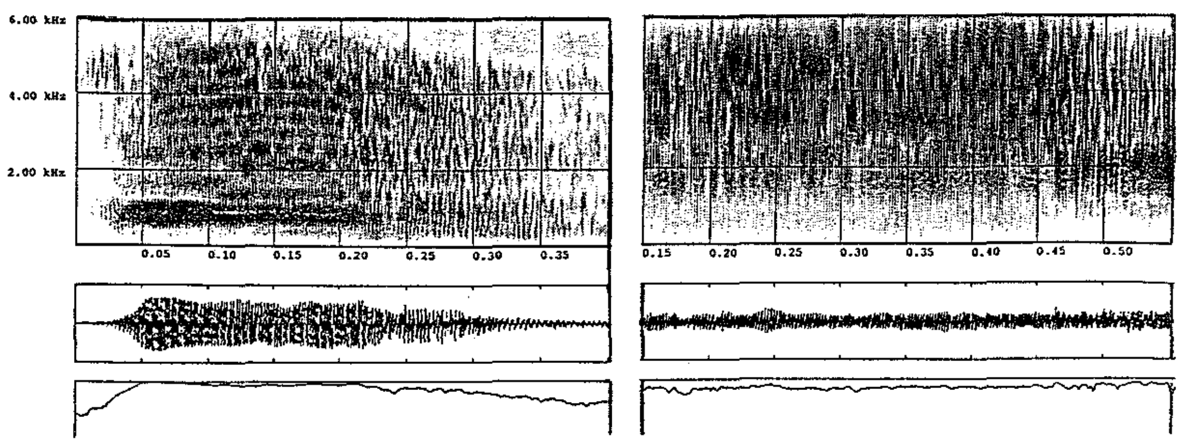
Figures 2a and 2b. Spectrograms of fully resonant (2a) and quasi-resonant (2b) sounds.

3.0 ANALYSES METHODS

Acoustic analyses were performed by three different methods, one based on the Fast Fourier Transform (FFT) and two on Linear Predictive Coding (LPC). We are not necessarily aiming at what method suits the analysis of (infant)speech best, but rather which methods are recommended in measuring infant utterances. This means that various measuring techniques could be used partly and/or simultaneously in order to obtain the most objective measures. On the next page we will give a brief description of the tested methods.

3.1 Spectrography

Although it is a time comsuming method the advantage of spectrography is that the courses of the parameters are made visible as bands of energy.
The digital spectrograph, implemented on the micro VAX mini-computer, performs a Fourier analysis on the speech signal as a function of both frequency and time. The density of the darkness on the plot corresponds to the amount of acoustic energy within the required range of frequencies. The bandwidth of the analysing filter is manipulated by means of a window length. This holds that a 4 msec window is required to approximate a bandwidth of 250 Hz. Every 1 msec the analysing window is shifted to yield a clear display of the course of the formants.
Of each utterance a narrowband (25 msec) as well as a broadband (4 msec) spectrogram was made. Narrowband spectrograms are generally used to estimate pitch as they yield a good frequency resolution, whereas broadband ones are suitable to estimate formant frequencies, as they show a poor frequency, yet a good time resolution. The appropriate bandwidth is a major problem for the higher-pitched voices as some of our broadband spectrograms were more like narrowband ones, yielding a harmonic structure (Fig. 3a). However, changing the bandwidth to 2 msec -as was done for those utterances yielding a harmonic structure- often resulted in a noisy spectrogram, the bandwidth being too wide now (Fig. 3b).



Figures 3a and 3b. Spectrograms analysed by a 4-msec (3a) and a 2-msec window (3b).

3.2 Linear predictive coding (LPC)

Linear predictive coding, an increasingly popular analysing technique in speech science, usually offers a fairly stable parametric representation of the speech signal. The basic idea behind this technique is that each speech sample can be represented as a linear combination of its past values together with the 'current' input value, this being necessary as the source signal is unknown. The prediction coefficients responsible for the estimations of the formant frequencies and bandwidths are determined by minimizing the mean-square error between the actual and the predicted values of the speech samples (Least square minimalisation). By solving the prediction coefficients an optimal fit to the envelope of the speech spectrum is found. Contrary to the robust LPC method (see 3.3) this conventional one weights all prediction residuals equally (Split-Levinson). As the algorithm provides an optimal fit every two poles will always yield a formant frequency estimate. In our data the 25-msec window was shifted every 10 msec.

3.3 Robust linear predictive coding (RLPC)

Based, in essence, on the same ground, the algorithm of the robust linear prediction method differs from the conventional LPC one in the relative weight of the error signal (Lee, 1988). Rather than minimizing the sum of squared residuals as is done by the conventional LPC method, RLPC minimises the sum of appropriately weighted residuals. As the non-Gaussian nature of the source is taken into account, outliers (occuring most frequently at the glottal opening) are deweighted. This holds that the prediction polynomial -from which the formants and bandwidths are estimated- will yield a smaller variance compared to the conventional LPC and therefore more accurate estimates. As with the conventional LPC method the analyses was performed by analysing the signal every 10 msec by a 25-msec window. We expect the robust method to estimate the prediction coefficients well as the method performed satisfying when tested on relatively high-pitched synthetic stimuli (Weenink, 1988).

4.0 PROCEDURE

The first two formant frequencies were estimated from 50-msec segments by one examiner from the wideband spectrograms. Dependent on the length of the utterances the estimation was repeated two or three times per spectrogram.
By tracing the 50-msec segments, in which the formant frequencies had been estimated in the spectrograms, in the LPC and RLPC analyzing files the formant frequencies of the three methods could be compared. As the formant frequencies by LPC and RLPC had been estimated in a 10 msec-frame every 50 msec sample contained 5 values. These were averaged to give one formant frequency estimate. Together there were approximately 1000 measurements (50 utterances * 2 (or 3) samples * 3 methods * 3 periods).
The fundamental frequency was estimated by the conventional LPC method and by adding and dividing 10 harmonics from the narrowband spectrograms. In table 1 the mean fundamental frequencies of the 9th, 12th, and 15th months as estimated from the spectrograms and by LPC are listed together with the frequency range (the lowest and highest F0 value).

Table 1. Mean fundamental frequency in Hertz and frequency range as estimated by narrowband spectrograms and by LPC.

	Spectrog.(Hz)	Freq. range(Hz)	LPC(Hz)	Freq. range(Hz)
9th month	405	270-1000	434	191-889
12th month	353	250-450	450	240-885
15th month	467	290-1000	486	270-909

4.1 Comparing formant frequency estimates

The estimations of formant frequencies differ considerably across methods and months. The formant frequencies of the months under test are plotted per analysing technique in Figure 4. It may seem as if the spectrographic plots are based on less values than the conventional ones; this is not the case though. Due to the relatively large frequency range (50 Hz) in which the formant frequencies are estimated many data points coincide, resulting in relatively few marks in the spectrographic plot.

Note that the plot belonging to the RLPC estimates is scaled differently from those of LPC and spectrograms. With RLPC the F2 estimate is often twice as high as those of the other two methods. The first and second formant frequencies of the conventional LPC and spectrographic methods range between 400 Hz-1800 Hz and 1400 Hz-3400 Hz, respectively, corresponding to the results of Kent and Murray (1982) and Kent et al. (1987).

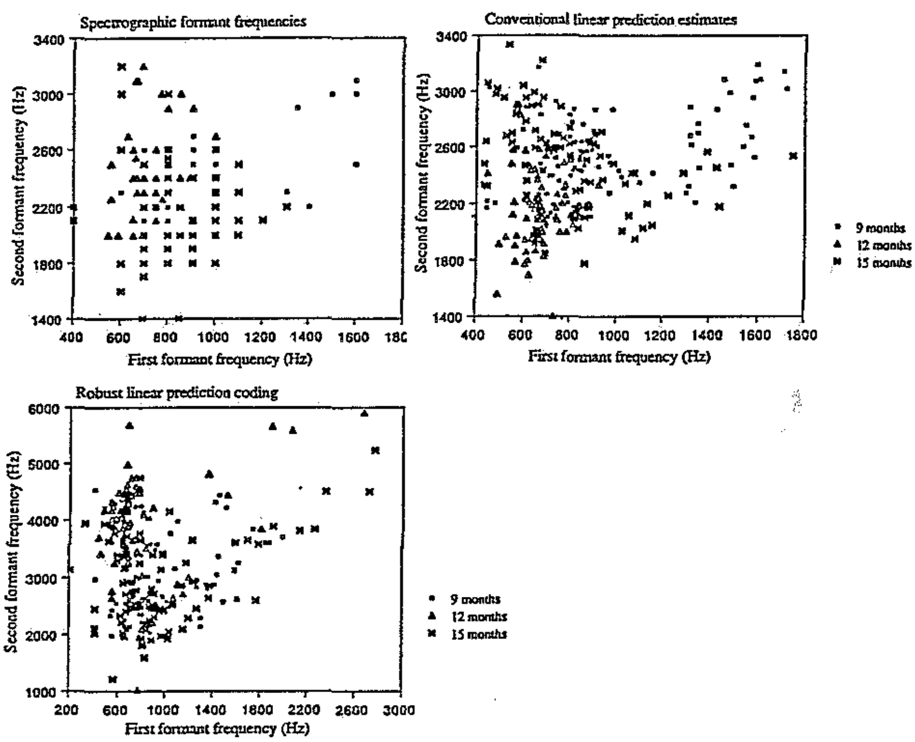
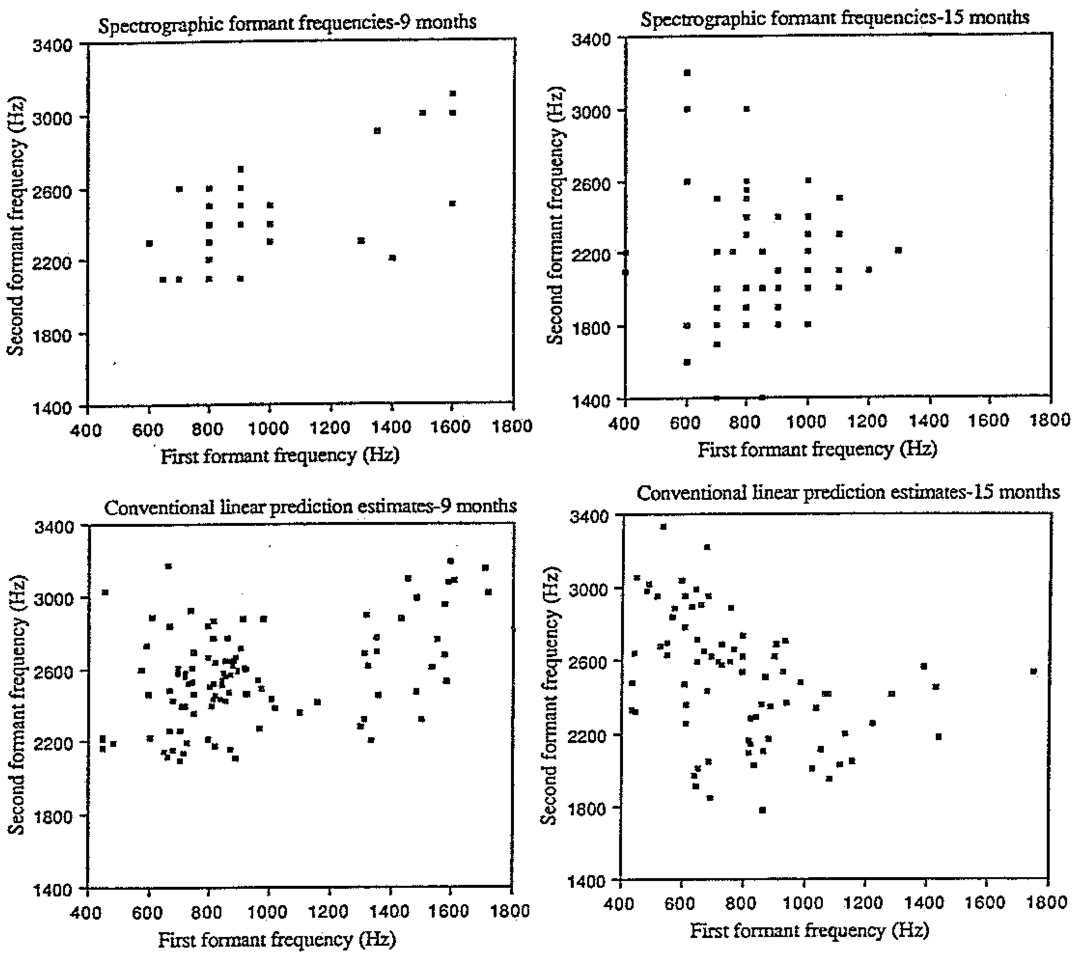


Figure 4. Formant frequency estimates by spectrography, the conventional LPC and the robust LPC methods. Every plot contains the estimates of the 9th, 12th, and 15th months. Note that the RLPC plot is scaled differently.

The formant frequencies of the 9th, 12th, and 15th month as measured by the conventional LPC and spectrogaphic method are also plotted separately in Figures 5a and 5b, respectively. By both methods the formants of the 9th month are estimated higher than those of the 12th and 15th months. In the latter two periods the formant frequencies tend to shift towards lower values. This is in agreement with the results of Kent et al. (1987) for two boys, almost as old as our infant. The results are interpreted "to reflect the lengthening of the vocal tract during this period" (p.67). The very high frequencies, especially in our 9-month data may be explained by the unarticulatory nature of the utterances. This finding corresponds once again to those of Kent et al. (1987) on the hearing-impaired boy at 8 months of age.



Figures 5a and 5b. Formant frequency estimates by spectrography (5a) and conventional LPC (5b) for the 9th and 15th month separately.

Apart from the central-like utterances being a result of the unexploited functioning of the vocal tract their frequencies may also be positioned at multiples of the fundamental. The relation between the fundamental frequency and the formant frequencies was therefore tested for the 9th and 15th month for both the conventional LPC and the spectrographic methods (the robust LPC method does not yield F0 data). In Figure 5C we have plotted F1 and F2 as multiples of the fundamental. In both methods many data points coincide or are very close to the exact multiple. As the estimates by the LPC method are decimally more accurate than those of the spectrographic one it expected that more exact multiples will occur in the latter.

In our data the F1 as measured by the spectrogams occurs more often at exact multiples in the 9th month than in the 15th one. In the 9th month data of this method the first formant occurs 27 times at 2.0 * the fundamental (and 10 times at 1.9 * the F0) versus 6 times at 2.0 * the F0 in the 15th month. However, the F1 of the LPC method yields 15 exact multiples at 2.0 * the fundamental at 9 months versus 14 multiples at 15 months (and 23 at 1.0 * F0). In general, the F2 is distributed over the entire frequency range, although more exact multiples occur in the spectrographic method. Also, the data of F2 at 15 months as estimated by LPC hardly yields any formants at multiples of the F0. It is not unthinkable that the formant frequencies of very young infants are positioned at multiples of the fundamental. In the perception of harmonic complexes it is found that, when discriminating between two high-pitched harmonic complexes, listeners are aided by a harmonic coinciding with a formant frequency (van Wieringen and Pols, this volume). Still, we should be careful with interpretations concerning the production of infants as these findings may, for the greater part, well be the result of the limitations of the analysing techniques. Future research will have to substantiate our findings.

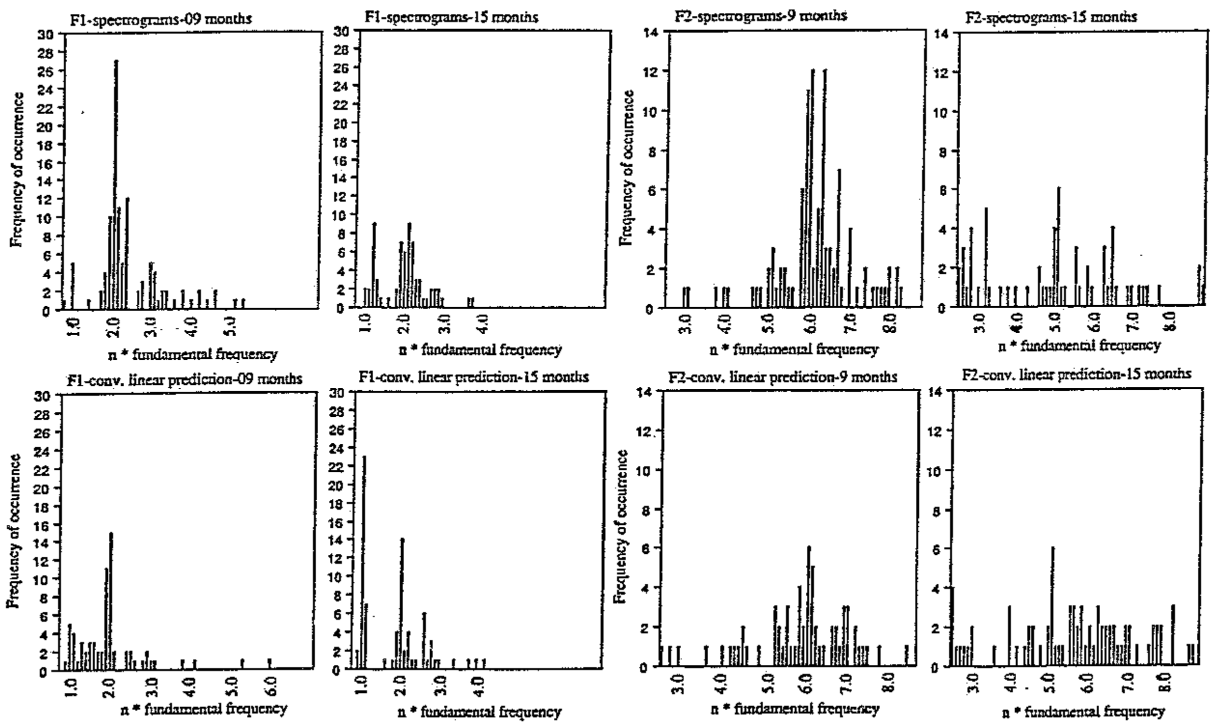


Figure 5C. Histograms of the number of times formant frequencies are estimated at multiples of the fundamental frequency. Note the different scales of the ordinates for F1 and F2.

Generally speaking, the formant frequencies can be read reasonably accurately from spectrograms. That is to say, if there is energy the frequency is presumed to be a formant. However, very often a broad band of energy is visible at a relatively low frequency. This band usually merges with the fundamental (see Figure 2b). As no energy is present as the higher frequencies one can debate on the lower frequency band. Does it feature the vocal development? Do infants have to learn to make fully resonant sounds? The F1 usually occurs at a multiple of the harmonics of the fundamental frequency. As the fundamental frequency is high -450 Hz on average- the energy present will be due to the widely spaced harmonics. Moreover, if the bandwidth of the analysing filters is not chosen appropriately the spectrogram yields a harmonic structure (Figure 3a).

In order to confirm the agreement within the two methods the formant frequencies of the conventional LPC method and those of the spectrgrams are plotted together in Figure 6. This was done only for the 9th and 15th months as the analysing techniques are our main concern and the 12th month did not provide us with extra information on the vocal development of the hearing-impired infant.

The LPC estimates resemble those of the spectrograms well. The correlation between the first formant frequencies as estimated by the conventional LPC technique and spectrograms is 0.93 and 0.61 for the 9th and the 15th months, respectively. The correlation of F2 for both months is 0.58 for the 9th and 0.73 for the 15th month. Due to 'missing' second formant frequencies in the 9th month, there are 110 first formant and 100 second formant samples. The plots for the 15th month include 97 samples. The relatively high correlations between these two methods indicate that both agree well, although it must be considered seriously whether the agreement is a result of the limitations of both methods or of the vocal development of the infant. For instance, the relatively high first formant cluster at 9 months (Figure 6) is also visible in the 9-month plots of Figures 5a and 5b. A similar cluster is absent in the 12th and 15th month plots.

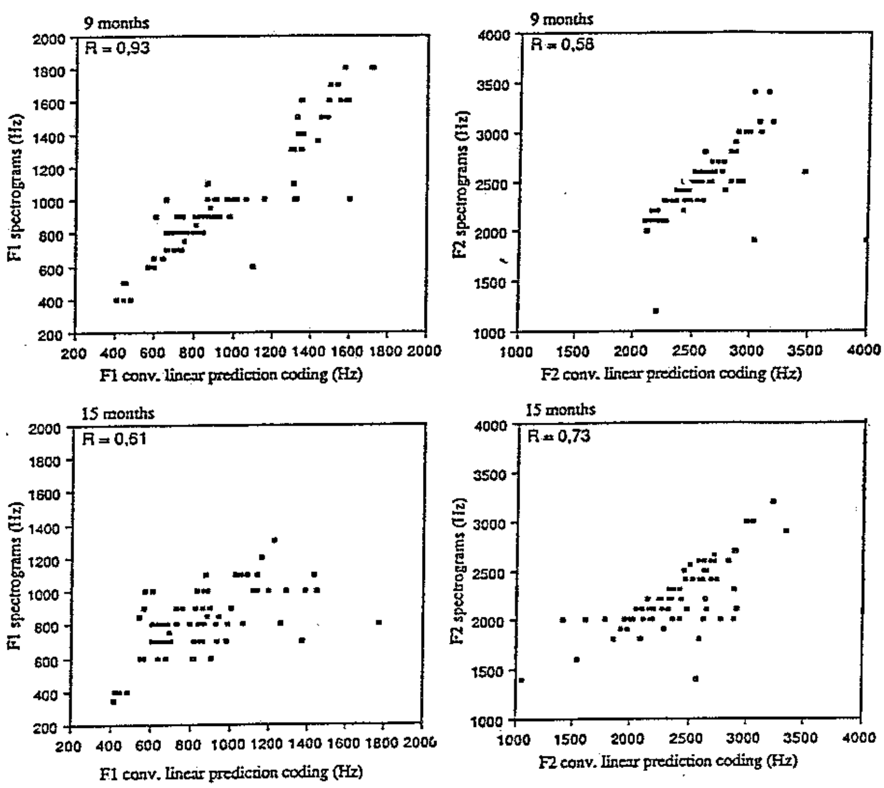


Figure 6. Correlation between the spectrographic and conventional LPC methods for the first and second formant frequency and 9th and 15th month separately.

5. CONCLUSIONS AND DISCUSSION

The robust linear prediction method does not suit our data quite well. Although this method is thought to separate the source and the filter better than conventional LPC it is not recommended for very young infants. The structure of the signal is in this respect a problem, 'quasi-periodic' periods often being very difficult to detect. Apparently, the conventional LPC and the spectrographic methods can cope better with the lack of periodicity. An alternative is to estimate the formant frequencies by a digital segment spectrograph (digitized method of Wempe, 1979). By means of this technique it is possible to calculate a spectrum on only one pitch period. If the period is accurately selected, formant frequencies can be estimated reliably. The segment spectrograph was tested on a few utterances and yielded the same formant frequencies for those utterances as measured by the spectrographic and conventional LPC methods. The results support our findings that the high fundamental frequency as well as the (lack of) structure of the speech signal are major problems in the spectral analyses.
If both broad- and narrowband spectrograms are used simultaneously, possibly combined with the digital segment spectrograph in order to compare estimates, results appear to be satisfying. The bandwidth of the filter is of extra importance in the broadband spectrograms. If the bandwidth of the filter is lower than the fundamental frequency, the spectrogram will display a harmonic-like structure. However, a very wide bandwidth causes a noisy spectrum: compare Figures 3a and 3b. In general a 600 Hz bandwidth is appropriate to extract formant frequencies for infants. The advantage of spectrograms lies in the algorithm (FFT), which is more straightforward than that of LPC.
As for the data obtained from the tested methods we have the impression that formant frequencies of the vocalic utterances of this infant differ essentially between the 9th and 15th month, possibly resulting from the fact that he started canonical babbling at the age of 14 months, which is in agreement with the results on hearing-impaired infants as given by Oller et Eilers (1988). Moreover, the 9th month data are quite in agreement with the formant data Kent et al. (1987) presented for the hearing-impaired twin boy at 8 months. In his and our data especially the two clusters of data points in the F1-F2 plots are striking, although at this stage it is not clear yet whether these results represent early (quasi-resonant) utterances characteristic for hearing impaired children, or whether they are merely inherent to the inaccurate or improper analysing techniques.
However, to obtain a proper frame of reference it will be necessary for each of the five speech developmental stages as mentioned in the introduction, to first analyse a number of vocalic utterances of normal hearing infants before we are able to characterise systematically the "more difficult to interpret utterances" within normal development as well as within disorders.

ACKNOWLEDGEMENTS

We would like to thank the infant's parents and Jeannette van der Stelt for making and supplying us with the audio-recordings, David Weenink for his assistance concerning the analysing techniques and Louis Pols for helpful comments and suggestions.

REFERENCES

Gilles, S., Koopmans-van Beinum, F.J., and Stelt, J.M. van der (1988). "The prelinguistic stage", in: P. Jordens & J. Lalleman (Eds.), Language Development, Foris Publications, Dordrecht-Holland, 29-56.

Kent, R.D. and Murray, A.D. (1982). "Acoustic features of infant vocalic utterances at 3, 6 and 9 months", J. Acoust. Soc. Am., 72 (2), 353-365.

Kent, R.D., Osberger, M.J., Netsel, R., and Goldschmidt Hustedde, C. (1987). "Phonetic development in identical twins differing in auditory function", J. Speech Hear. Disord, 52, 64-75.

Koopmans-van Beinum, F.J. and Stelt, J.M. van der (1979). "Early stages in infant speech development", Proc. of the Institute of Phonetic Sciences Amsterdam, 5, 30-43.

Koopmans-van Beinum, F.J. and Stelt, J.M. van der (1986). "Early stages in the development of speech movements", in: B. Lindblom & R. Zetterström (Eds.), Precursors of Early Speech, Macmillan Press Ltd., England, 37-50.

Lee, C. (1988). "On robust linear prediction of speech", IEEE Trans. ASSP. 36(5), 642-649.

Makhoul, J. (1975). "Linear prediction: a tutorial review", Proc. IEEE 63, 561-580.

Oller, D.K., Eilers, R.E., Bull, D.H. and Carney, A.E. (1985). "Prespeech vocalizations of a deaf infant: a comparison with normal metaphonological development", J. Speech Hear. Res. 28, 47-63.

Oller, D.K. and Eilers, R. E. (1988). "The role of audition in infant babbling". Child development, 59, 441-449.

Robb, M.P. and Saxman, J.H. (1985). "Developmental trends in vocal fundamental frequency of young children", J. Speech Hear. Res. 28, 421-427.

Robb, M.P.and Saxman, J.H. (1988). "Acoustic observations in young childeren's non-cry vocalisations", J. Acoust. Soc. Am., 83(5), 1876-1882.

Weenink, D.J.M. (1987)." Robust linear prediction analysis of vowels", Proc. of the Institute of Phonetic Sciences Amsterdam, 11, 79-88.

Wempe, A.G. (1979). An experimental segment spectrograph based on some notes on frequency analysis of speech segments, Proc. of the Institute of Phonetic Sciences Amsterdam 5, 44-102.

Wieringen, A. van & Pols, L.C.W. (1989). "Discrimination in densely and sparsely spectralled harmonic complexes", Proc. of the Institute of Phonetic Sciences Amsterdam, 13, 27-43 (this volume).