

# Infant Communication: Cry and Early Speech

Thomas Murry

*Speech Pathologist, Veterans Administration  
Medical Center; and Adjunct Assistant  
Professor of Surgery, University of California,  
San Diego, California*

and

Joan Murry

*University of San Diego*



College-Hill Press • Houston, Texas

fare.

usual features  
unexpectedly.  
2. Anatomy and  
Infant Death  
3. Washington:  
and Welfare.

apnea and the  
clinical and  
6-54.

apnea and the  
Presented at  
Sudden Infant  
Infants at Risk  
to Prevention,

5. Apnea and  
and sleep.

sleep apnea and  
nation study.

statistical theory  
lications. J.

possible cause  
on. Canadian  
nt Death, pp.

J., Partanen,  
ant Cry: A  
Clinics in  
on: Spastics

# 9

## CHAPTER

### Patterns of Fundamental Frequency and Vocal Registers

Patricia Keating

#### INTRODUCTION

This chapter reports observations on vocal register use and register shifts in young children. It also reports on the fundamental frequency ( $f_0$ ) values observed in the speech of four children before the appearance of their first words. The goal of this chapter is to add to our knowledge about normal children's speech vocalizations and ultimately to develop a model of the acquisition of prosodic features.

There are few sources of data on the range of  $f_0$  in young children's speech. Studies reporting average  $f_0$  for reading include Peterson and Barney (1952), Fairbanks, Wiley, and Lassman (1949), and Fairbanks, Herbert, and Hammond (1949); these averages vary from about 250 to 300 Hz. Other studies, such as Sheppard and Lane (1968), report infant  $f_0$  for continuous vocalization, including cries. Average  $f_0$  for 95-sec samples of two infants over a five-month period varied from 398 to 438 Hz. Still other studies, such as Fairbanks (1942), Prescott (1975), and Murry, Amundson, and Hollien (1977), treat only infant cries. In such studies, average  $f_0$  typically ranges from 410 to 585 Hz. Keating and Buhr (1978) provide information on  $f_0$  and vocal register use in the speech of children ages 33 to 169 weeks. In their study,  $f_0$  ranged from a low of 30 Hz to a high of 2500 Hz. Each sample for each child had a  $f_0$  range of at least several hundred Hz, with some children showing extreme  $f_0$  variation. No longitudinal trends were apparent in the range of  $f_0$  used. In addition, utterances were analyzed as

containing one or more of three registers: fry, modal, and high. (See below for descriptions of these registers.) In fry register,  $f_0$  values ranged from 30 to 250 Hz; in modal register, from 150 to 700 Hz; and in high, from 380 to 2500 Hz. These values cover a much wider range than had previously been reported for child speech.

The present chapter, using the methodology of Keating and Buhr (1978), offers a number of advantages for the study of speech acquisition. First, this methodology excludes  $f_0$  values of cries and vegetative sounds (e.g., spitting and burping) which provide data on the inherent properties of the larynx but are of unknown relevance to speech. Second, the data are derived from spontaneous speech recorded in a natural, familiar setting over an extended period of time. This is particularly true in the present study. Since these children were recorded from such an early age, recording sessions became a normal part of life for them. Third, the data are presented as overall ranges as well as averages. Fourth, the data are classified by vocal register, giving some indication of longitudinal trends in register use.

#### METHODOLOGY

The data reported are drawn from tape recordings of child speech in the Brown Phonetics Laboratory, directed by Philip Lieberman. These recordings are of several children playing at home, usually with their mothers, and were recorded approximately every two weeks using a Nagra 4.2 tape recorder and either an AKG or Sony ECM-50PS microphone. Ages range from ten days to over five years in this corpus; substantial longitudinal data are available, although no single child has been recorded over this entire period of time.

The samples were chosen to span at least two of three speech acquisition stages, which are described below. Such longitudinal data were available for four children. Table 1 includes a listing of the samples. It should be noted that the third sample for child JD, at 53 weeks, was the last before clearly recognizable words appeared; child LS did not produce such words for some time after the session at 66 weeks; the sample for child RC at 69 weeks is the last in her corpus.

There are clear phonetic differences characterizing

... fry, modal, ons of these ranged from 30 to 700 Hz; and values cover a n reported for

ogy of Keating ntages for the is methodology tative sounds de data on the re of unknown e derived from ral, familiar ime. This is . Since these in early age, t of life for overall ranges re classified idication of

recordings of Laboratory, ordings are of ly with their ely every two either an AKG from ten days substantial ough no single ire period of

it two of three cribed below. ble for four f the samples. for child JD, y recognizable such words for the sample for orpus. characterizing

Table 1. Overall  $f_0$  range for each sample for each child.

Child	Sex	Ages (weeks)	$f_0$ Ranges
JD	F	16	30 - 600
		33	80 - 2500
		53	? - 900
LS	M	16	15 - 1250
		33	25 - 1000
		66	55 - 1150
GM	M	16	20 - 1440
		36	25 - 2750
RC	F	33	80 - 500
		69	55 - 900

the three stages for these infants. Before 12 weeks, there are few speech-like sounds during any one session. By 16 weeks, breathy vowel-like sounds [ae, ε, a] begin to occur loudly and often enough for analysis. These vowels occur alone, in groups of phonetically similar vowels, or in syllables with glottal stops or [h]. After 20 weeks, longer strings of various vowels occur. During this period, CV syllables also begin to appear, but they do not occur consistently until after 40 weeks. It is these long sequences of syllables with consonants such as nasals and stops which emerge in the period around 40 weeks and that are typical of what is called babbling. After this period, until the use of recognizable words, these babbled utterances develop so that they more closely correspond to possible English words. In particular, they shorten to one- or two-syllable forms, sometimes with glides or final consonants.

Samples were chosen for each child at 16 and 33 weeks, where possible, and prior to recognizable words. A sample consists of the first 50 vocalizations in each session, with some exceptions.

The following types of vocalizations were eliminated: cries or vegetative noises, utterances with interfering noise in the background, low-volume utterances, ingressive and voiceless utterances. The end of an utterance was taken to be the point at which the child took a breath. This point was generally easy to determine by listening to the tape since our recordings are made by holding a sensitive directional microphone within a few inches of the child's mouth, thereby emphasizing breathy noise.

Spectrograms were made on a Kay Songraph, model 6061B. The input was recorded at high speed (160-16,000 Hz), so that there was an effective narrow bandwidth of 90 Hz and an effective wide-bandwidth of 600 Hz. Various settings on the 6076C scale magnifier were used as appropriate; setting the upper limit to 50% displays 8000 Hz on the spectrograms, while setting it to 40% displays 6400 Hz. Fundamental frequency was measured from narrow band spectrograms at the fifth harmonic (or the next highest one possible if the fifth was absent or obscured) to the nearest 25 Hz. Values below about 200 Hz were determined by counting striations per unit of time on wide band spectrograms.

#### DATA

Table 1 lists the samples available for each child and the total  $f_0$  range for each sample. In several cases, this range is quite wide. Table 2 divides  $f_0$  values in each utterance into three vocal registers: fry, modal, and high. The lowest and highest  $f_0$  values within each register were noted for each utterance. The range within each register is presented, as well as the mean low and high values and their standard deviations. Assignment of an utterance, or portion of an utterance, to a register was conservative; it would be considered "modal" unless there was a clear indication to the contrary. Thus, there may actually be more instances of vocal fry and high register use than are tabulated here.

The lower  $f_0$  values are found in fry register, also known as "creaky voice" (Ladefoged 1971; Hollien 1974). Fry register is associated with widely and irregularly spaced striations on wide band spectrograms. Generally, fry register for this age group occurs at  $f_0$  values below approximately 250 Hz. Often, a few fry pulses occur at the onset of phonation for about 50 or

re eliminated:  
 th interfering,  
 e utterances,  
 The end of an  
 igh the child  
 ally easy to  
 our recordings  
 al microphone  
 uth, thereby  
 graph, model  
 i speed (160-  
 ctive narrow  
 -bandwidth of  
 :ale magnifier  
 upper limit to  
 grams, while  
 Fundamental  
 pectrograms at  
 e possible if  
 nearest 25 Hz.  
 d by counting  
 spectrograms.

or each child  
 . In several  
 2 divides  $f_0$   
 al registers:  
 best  $f_0$  values  
 ch utterance.  
 d, as well as  
 eir standard  
 or portion of  
 ive; it would  
 was a clear  
 may actually  
 register use

egister, also  
 Hollien 1974).  
 and irregularly  
 pectrograms.  
 ? occurs at  $f_0$   
 en, a few fry  
 or about 50 or

Table 2.  $f_0$  range (in Hz), mean lowest value, and mean highest value for each vocal register for each sample (child and age in weeks).

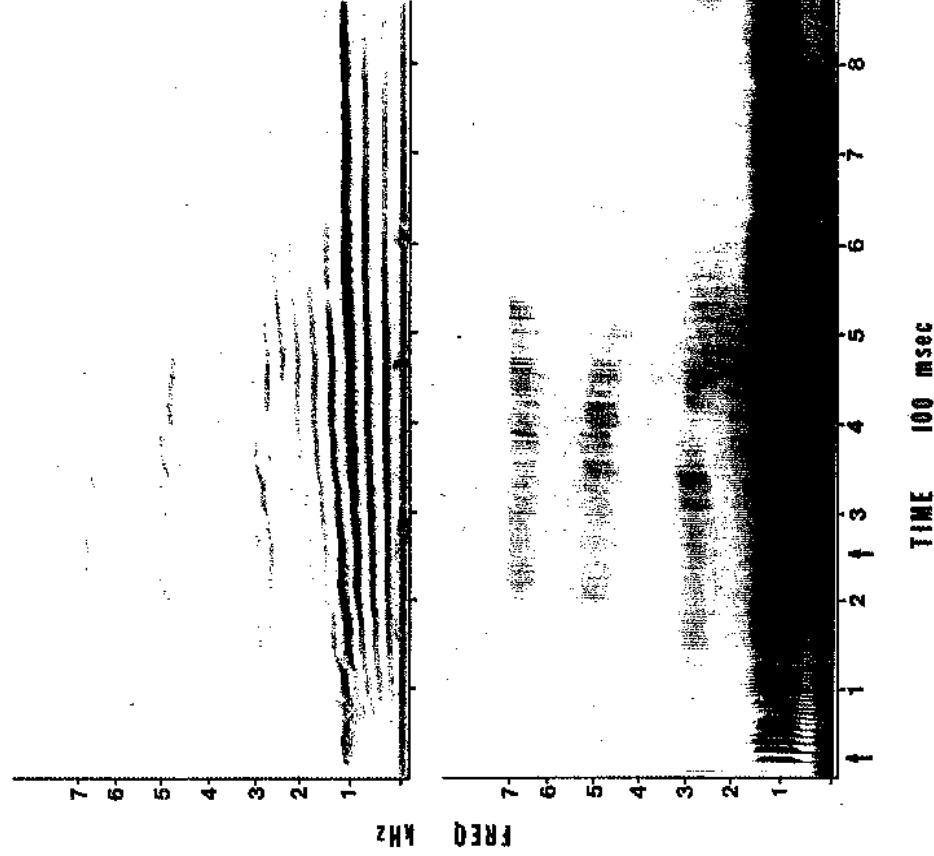
Sample & Register	N	Overall Range	Range of Lows	Mean Low (+ S.D.)	Range of Highs	Mean High (+ S.D.)
JD - 16						
FRY	12	30-290	30-135	73 (37)	165-290	215 (38)
MODAL	50	150-415	150-300	242 (37)	225-415	299 (50)
HIGH	4	400-600	400-500	463 (48)	475-600	534 (57)
JD - 33						
FRY	6	80-330	80-110	94 (13)	240-330	293 (34)
MODAL	48	150-575	150-425	315 (60)	300-575	424 (70)
HIGH	13	350-2500	350-1000	545 (178)	500-2500	990 (621)
JD - 53						
FRY	1	?	?	?	?	?
MODAL	50	200-600	200-325	281 (34)	275-600	380 (68)
HIGH	6	400-900	400-550	467 (61)	450-900	625 (175)
LS - 16						
FRY	40	15-300	15-160	72 (45)	160-300	219 (41)
MODAL	50	175-575	175-350	251 (46)	300-575	372 (43)
HIGH	11	500-1250	500-850	640 (98)	540-1250	835 (213)
LS - 33						
FRY	24	25-340	25-215	103 (51)	160-340	267 (50)
MODAL	50	200-490	200-400	303 (43)	300-490	382 (38)
HIGH	6	400-1000	400-750	549 (169)	420-1000	677 (260)
LS - 66						
FRY	10	55-340	55-180	127 (46)	160-340	253 (66)
MODAL	45	220-575	220-450	323 (46)	300-575	359 (53)
HIGH	5	500-1150	500-940	698 (172)	500-1150	777 (246)
GM - 16						
FRY	16	20-340	20-200	94 (57)	140-340	219 (54)
MODAL	49	150-575	150-400	276 (63)	275-575	388 (68)
HIGH	6	450-1440	450-790	622 (143)	575-1440	878 (297)
GM - 36						
FRY	16	25-290	25-160	96 (44)	140-290	199 (59)
MODAL	40	150-500	150-425	260 (85)	250-500	365 (60)
HIGH	16	350-2750	350-2000	645 (387)	500-2750	907 (540)
RC - 33						
FRY	8	80-290	80-160	104 (30)	120-290	216 (66)
MODAL	48	120-500	120-450	280 (57)	250-500	388 (52)
HIGH	0	-	-	-	-	-
RC - 69						
FRY	9	55-290	55-225	116 (62)	160-290	236 (43)
MODAL	50	200-650	200-450	305 (71)	250-650	434 (93)
HIGH	5	350-900	350-900	545 (244)	375-900	570 (239)

100 msec. Less often, an entire utterance will move in and out of fry register. At any given age, children differ in how much fry they use, and a child may use different amounts of fry at different ages. For this reason it is difficult to make statements about trends in the use of fry during speech development. The implications of this data for speech acquisition are discussed below.

It should be noted that the measurement of "highest fry" and "lowest modal" values is problematical since one register often changes smoothly into the other. If no obvious shift was evident, the lowest modal value

was taken to be the lowest one measurable from the narrow-band spectrogram, and the highest fry value was set arbitrarily at 10 Hz below that. Computer waveform displays of fry and modal phonation show the continuous pattern of modal phonations, even at low  $f_0$  values, versus the interrupted pattern of fry phonations. For example, Figure 1a shows narrow and wide band spectrograms of an utterance containing both fry and modal registers. Figure 1b shows the waveform of the utterance in fry register, and Figure 1c shows the waveform of the same utterance in modal register at approximately the points indicated in Figure 1a. Presumably, some criterion to differentiate the two types of waveforms could be devised, possibly on the basis of the pause between glottal pulses in fry

Figure 1a. Narrow- and wide-band spectrograms of utterance by RC at 33 weeks of [e], containing both fry and modal registers.



able from the fry value was outer waveform the continuous low  $f_0$  values, and both fry and waveform of the lc shows the l register at n Figure la. tiate the two ssibly on the ulses in fry

erance by RC at fry and modal

Figure 1b. Waveform from first arrow in 1a (16 msec); 1 point = .05 msec.

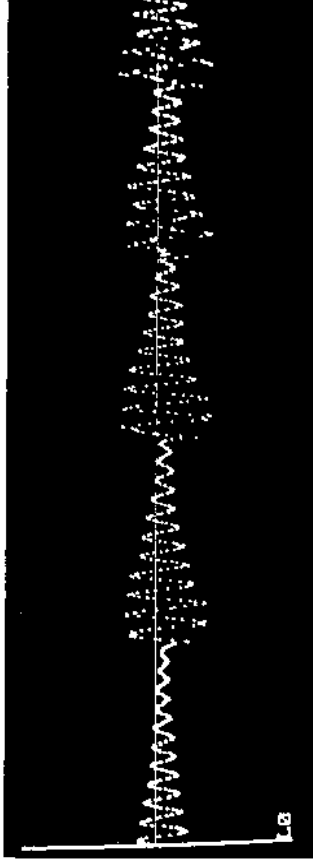
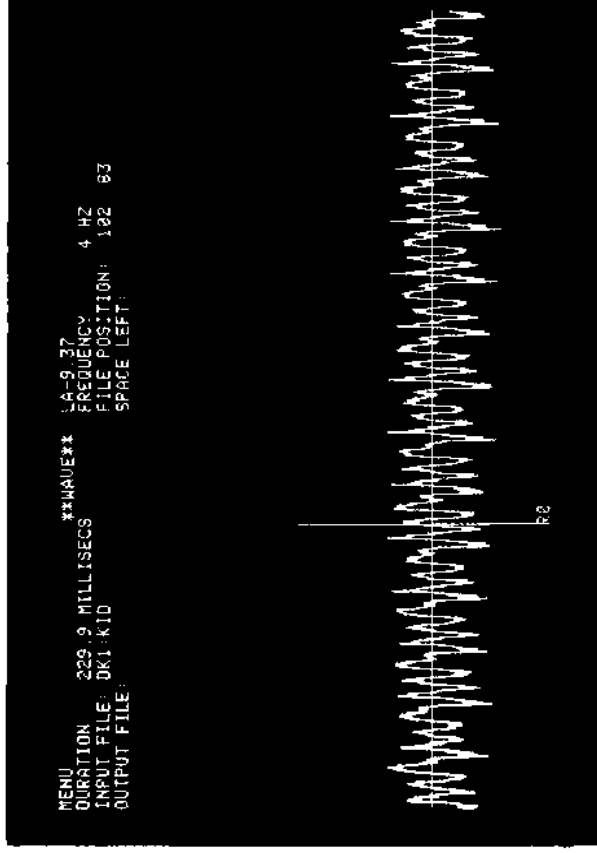


Figure 1c. Waveform from second arrow in 1a (246 msec).



register. Another criterion might be the spectra in each register. Figures 1d and 1e show the Discrete Fourier Transforms of the portions of the utterance shown in Figures 1b and 1c, respectively. In addition, the sophisticated computer editing techniques available on the PDP-11/34 allow a waveform to be arbitrarily segmented and individual segments listened to. With practice, such segmentation of a few pitch periods might permit perceptual categorization of individual portions of an utterance.

Utterances containing modal register comprise the bulk of the samples for each child. The range of  $f_0$  values in this register alone is wider than that previously thought to characterize non-cry



Figure 1d. Discrete Fourier Transform (DFT) corresponding to 1b; Hanning window is 25.6 msec wide with center at 28 msec.

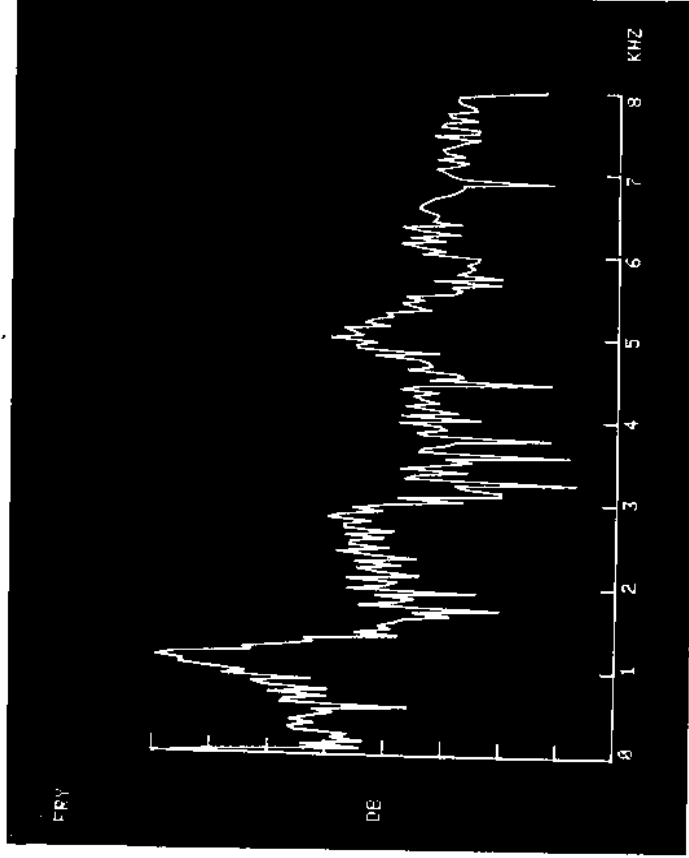
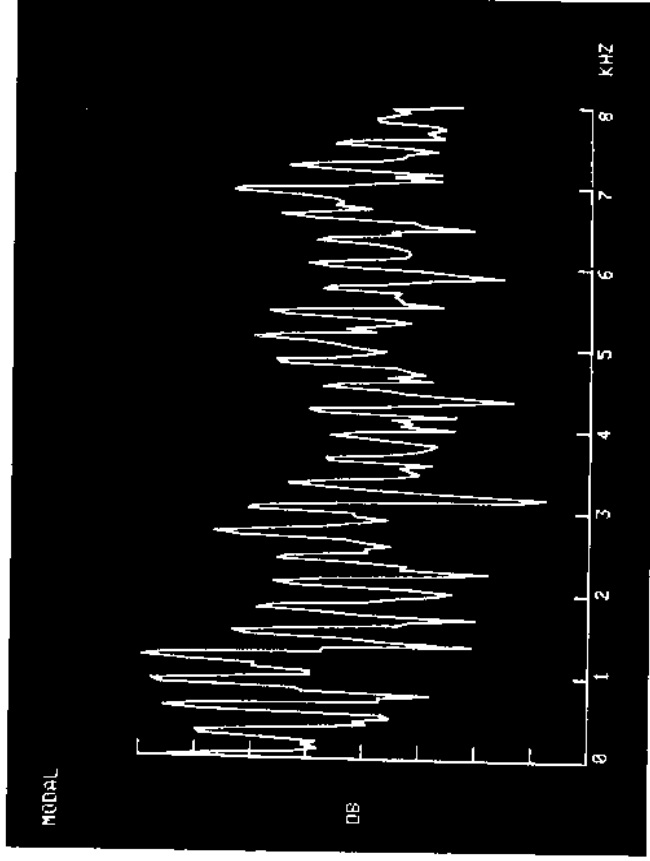


Figure 1e. DFT corresponding to 1c, with center at 258 msec.



ponding to lb;  
ter at 28 msec.



158 msec.



vocalizations. In order to give some indication of the frequency of occurrence of the higher  $f_0$  values in modal register, Table 3 presents a breakdown of the peak modal  $f_0$  values. The highest modal value of each utterance containing modal register was put into one of six  $f_0$  ranges.

"High" register is not to be confused with these higher modal values. Our term "high" is probably equivalent to "falsetto," but that term has been avoided since it has connotations of a false voice quality, which these children do not seem to be using. Additionally, it is unclear exactly how many registers are involved in these high  $f_0$  utterances; therefore, a single neutral term is preferred. It is quite likely that the very high and relatively uncommon  $f_0$  values included in "high" register are actually due to some fourth register, as their perceptual quality is somewhat shriller and more piercing. In general, high

Table 3. Breakdown of peak modal  $f_0$  values -- No. per  $f_0$  range.

Sample	200-99	300-99	400-99	500-99	600-99
JD - 16	25	19	6	-	-
33	-	14	26	7	1
53	1	32	14	2	1
LS - 16	-	39	9	2	-
33	-	29	21	-	-
66	-	35	9	1	-
GM - 16	1	23	21	4	-
36	1	27	10	2	-
RC - 33	1	19	26	2	-
69	2	14	18	12	4

register is characterized by high pitch and a thin quality. High register is usually associated with a discontinuity (or "shift") in the  $f_0$  contour on narrow band spectrograms, although this is not always the case. Additional shifts are sometimes seen at the higher  $f_0$  values which may then be further evidence of a fourth vocal register.

#### DISCUSSION

A number of interesting phenomena that can be observed in the spectrograms should be mentioned. In a very few cases, every  $n^{\text{th}}$  (for example, fifth) harmonic will be much darker than the others. This pattern seems to be the result of the child's having her vocal folds separated enough for that end of the vocal tract to be considered open. With the mouth also open, the formant frequencies generated will be spaced at equal intervals, and so individual evenly-spaced harmonics near these formants will appear darkened.

Another result of the vocal folds not being brought close together is the breathiness of much of the speech in these samples. When the vocal folds do not come together completely as they vibrate, air can pass through. In addition, the aerodynamic conditions that allow phonation to continue are affected such that the  $f_0$  is rather low, but there are few harmonics in the glottal source spectrum. In this case only the first one or two formants are periodically excited; higher formants are noise-excited.

In a few instances, what appears to be rapidly alternating fry and high register is seen on spectrograms. That is, in between the glottal pulses associated with fry, there are harmonics at multiples of about 1000 Hz. This  $f_0$  may be vibration of the edges of the vocal folds coupled to the supralaryngeal vocal tract at the first formant frequency. These edge vibrations would generate their own harmonics, even as the vocal folds as a whole vibrate in fry register. The slackness of the folds in fry may facilitate independent edge vibration.<sup>1</sup>

Usually fry and high register occur in utterances which also contain modal register phonation. There is generally a noticeable shift between modal and high register on spectrograms which may, in fact, be

<sup>1</sup>P. Lieberman: personal communication.

h and a thin  
iated with a  
our on narrow  
t always the  
seen at the  
ar evidence of

that can be  
be mentioned.  
ample, fifth)  
others. This  
hild's having  
at end of the  
the mouth also  
will be spaced  
evenly-spaced  
darkened.

being brought  
of the speech  
s do not come  
air can pass  
nditions that  
such that the  
monics in the  
nly the first  
cited; higher

o be rapidly  
is seen on  
flottal pulses  
at multiples  
ration of the  
supralaryngeal  
Y. These edge  
onics, even as  
fry register.  
y facilitate

in utterances  
ion. There is  
odal and high  
in fact, be

considered one criterion for the identification of high register in infants' speech. It is less obvious exactly where a shift between modal and fry registers occurs, although one register is clearly differentiated from the other. One indication of a register shift seems to be harmonic 'doubling', described by Buhr and Keating (1977).<sup>2</sup> In the case of doubling, extra harmonics appear on the spectrograms, evenly spaced between the harmonics of the fundamental and often fading in and out. The perceived pitch does not change, but the voice quality is quite distinctive, being somewhat raw and harsh. Although this quality is similar to that of cries or shrieks, these utterances are neither. In older children the doubling occurs during normal word and sentence production.

Table 4 indicates how often and in what registers doubling occurs. Only clear instances of doubling are tabulated, so the actual rate may be slightly higher. Both the upper and lower ranges of modal register seem to be characterized by doubling, especially in conjunction with register shifts. High register also shows the phenomenon in the absence of any apparent register shift. Figure 2a shows the narrow-band spectrogram of an utterance in high register with extended doubling. Figure 2b shows the Fourier Transform centered at the first arrow; the  $f_0$  is clearly about 900 Hz. Figure 2c shows the Fourier Transform centered at the second arrow. The harmonics through 5 kHz are multiples of 500 Hz, but there is no fundamental actually at 500. Figure 2d shows the

Table 4. Number of utterances per sample showing doubling, and in which register or between which registers each instance occurs.

Sample	Total	Fry-Modal	Modal	Modal-High	High
JD - 16	13	-	13	-	-
33	8	3	5	-	-
53	2	-	1	1	-
LS - 16	10	7	3	-	-
33	23	7	14	3	1
66	4	-	3	-	1
GM - 16	5	1	3	-	1
36	15	1	9	1	4
RC - 33	6	-	6	-	-
69	6	-	6	-	-

<sup>2</sup>'Doubling' appears to be the same phenomenon as the 'diplophonia' described in Mønsen (1979) and Mønsen, Engerbretson, and Vemala (1979).

Figure 2e

waveform of this portion. The pitch period is 1 msec, which corresponds to a  $f_0$  of 1000 Hz. A few msec later, however, there is an abrupt change in pitch period to 2 msec, which corresponds to a  $f_0$  of 500 Hz. Figure 2e shows the Fourier Transform centered at the third arrow of 2a with a weak fundamental at 500 Hz. A few msec later, the Fourier Transform shows a clear 500 Hz fundamental (Figure 2f). Later in the utterance, the process repeats itself in reverse until by 465 msec the  $f_0$  is again clearly 1000 Hz. The doubling phenomenon will be discussed further in the next section.

The use of wide  $f_0$  ranges, multiple vocal registers, and even register shifts involving doubling by the four normal infants of this study indicates that care should be taken in interpreting specific acoustic attributes of the vocal production of impaired children. High  $f_0$  has been observed in the cries of deaf (Jones 1965) and retarded (Ostwald, Phibbs, and Fox 1968) infants. In

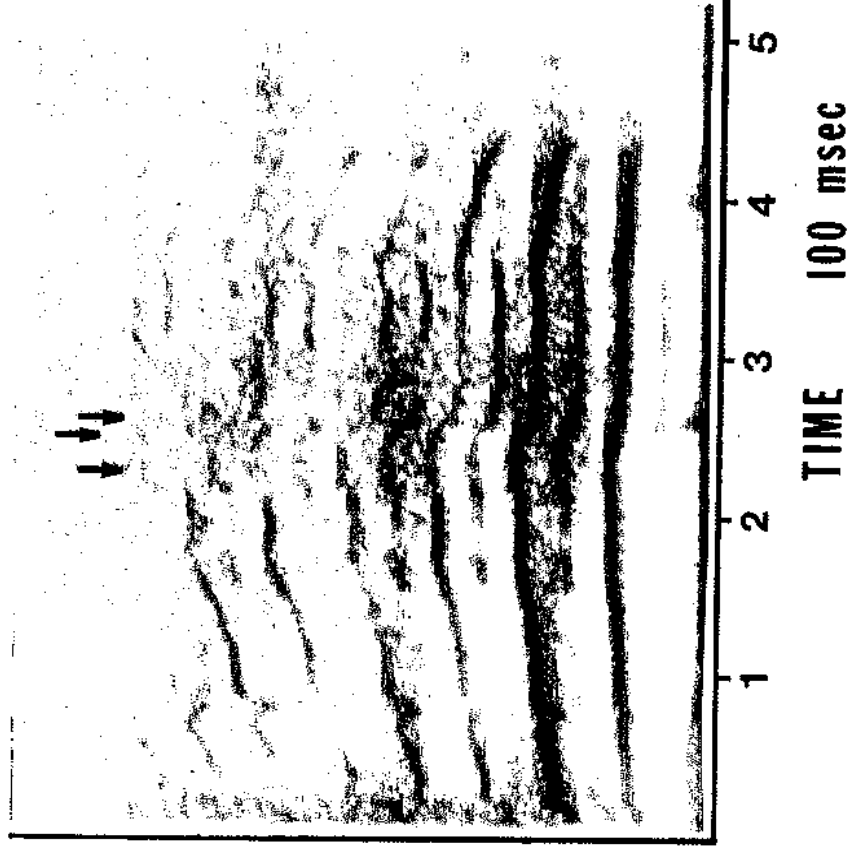
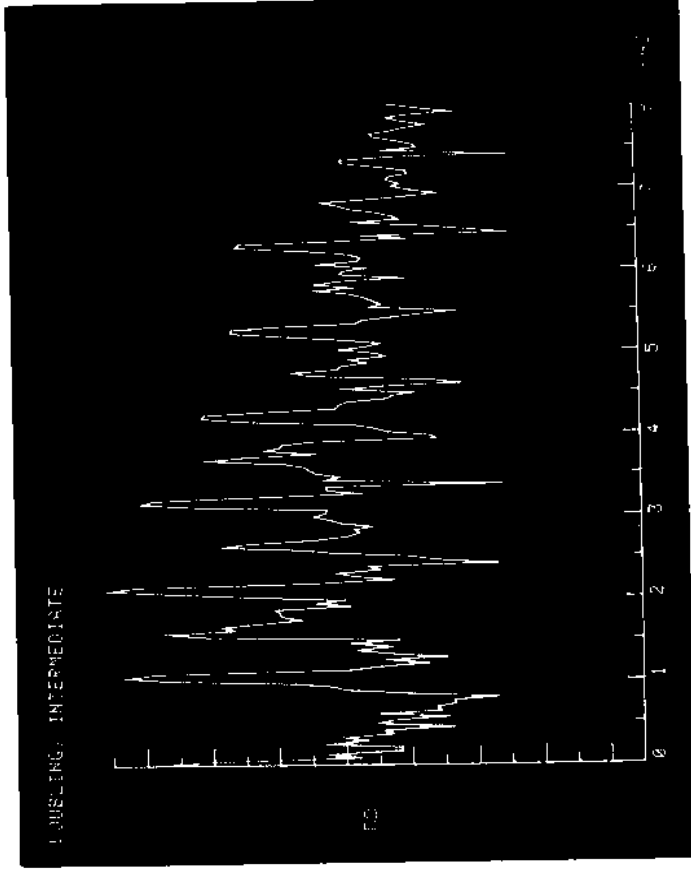
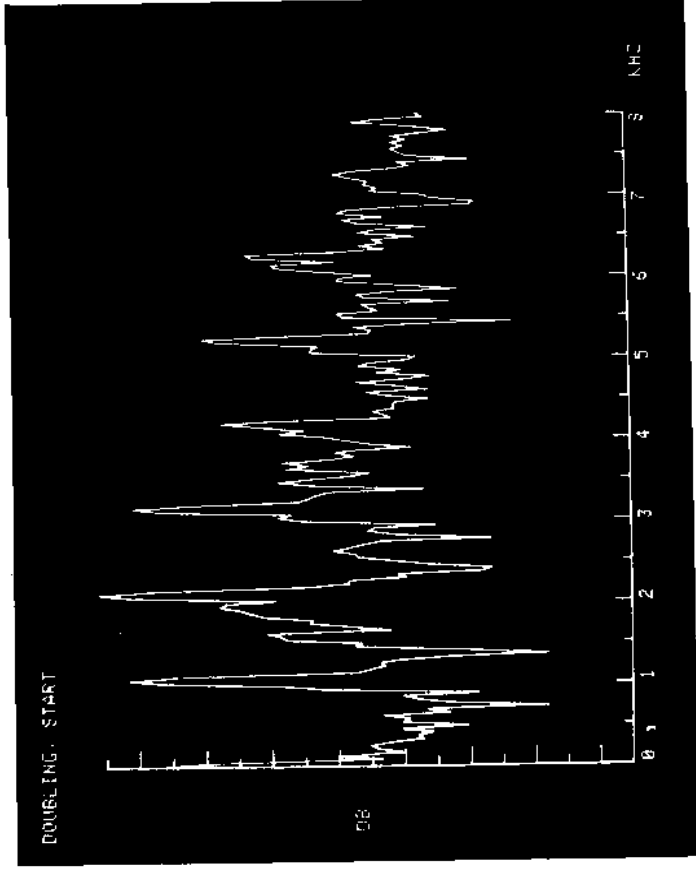


Figure 2a. Narrow-band spectrogram of utterance by GM at 36 weeks of [ae], showing extended doubling in high register.

is 1 msec, a few msec in pitch of 500 Hz. centered at the 500 Hz. A clear 500 utterance, by 465 msec the doubling in the next registers, by the four care should attributes an. High f<sub>0</sub> es 1965) and infants. In

Figure 2b. DFT centered at first arrow in 2a (230 msec).

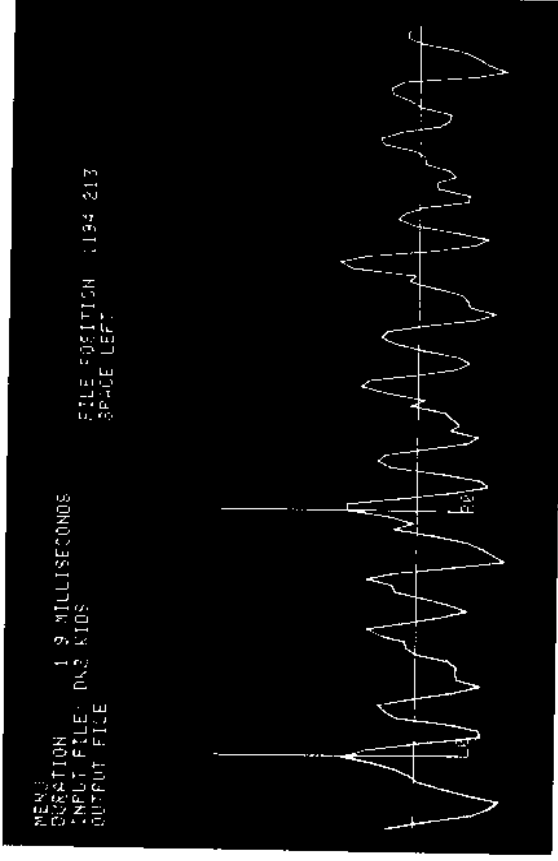


nsec

GM at 36 weeks  
h register.

Figure 2c. DFT centered at second arrow in 2a (250 msec).

Figure 2d. Waveform corresponding to 2c.



the latter study, spectrograms of supposedly "abnormal" cries show the sudden pitch changes characteristic of the register shifts of normal infant vocalizations. Also showing these shifts are spectrograms of "hyperphonation" in Truby and Lind (1965). Hyperphonation is their term for the very high  $f_0$  values, breathy phonation, and prominent formant 1 with its own harmonics that sometimes occur in cries. Truby and Lind also mention abrupt  $f_0$  shifts, suggesting that they are unique to newborn cries. The data presented here, however, indicate that all of these features are found in infant speech.

In particular, breathy phonation and abrupt  $f_0$  shifts are common for children as old as 169 weeks (Keating and Buhr 1978). Stark and Nathanson (1975) reported a greater incidence of high  $f_0$  in the cries of an infant who later died of Sudden Infant Death Syndrome than in those of normal infants. Again, the data presented here indicate the need for caution in making such connections. Neither fry register nor extremely high  $f_0$  values per se necessarily indicate the presence of any pathological condition or anomaly. The study of variation in normal infant speech production leads to the conclusion that a wide array of vocal output must be considered normal in speech development.



y "abnormal"  
characteristic of  
realizations.  
programs of  
and (1965).  
very high  $f_0$   
remant l with  
aries. Truby  
gesting that  
:a presented  
features are

apt  $f_0$  shifts  
eks (Keating  
) reported a  
of an infant  
lrome than in  
:a presented  
making such  
remely high  
presence of  
The study of  
ion leads to  
output must

Figure 2e. DFT centered at third arrow in 2a (258 msec).

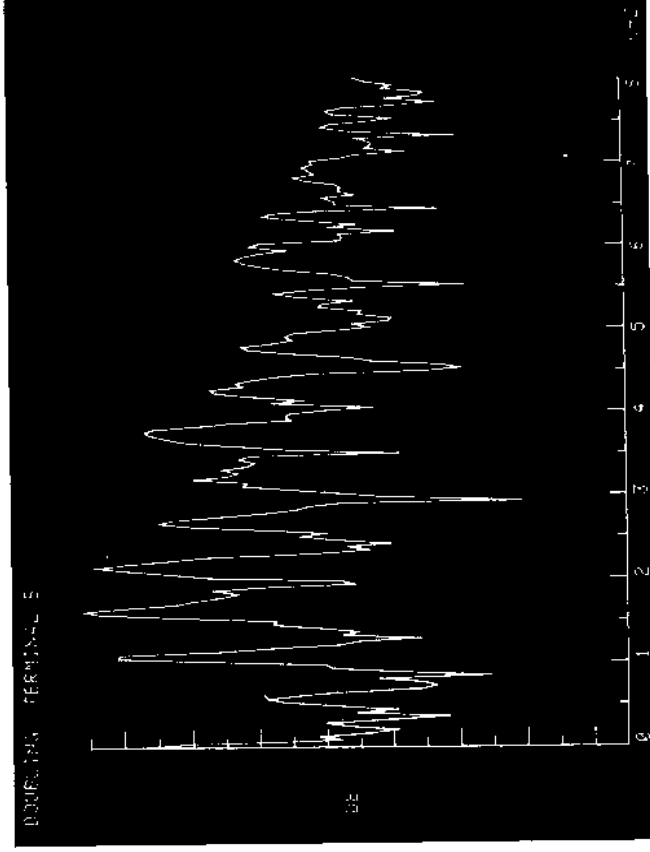
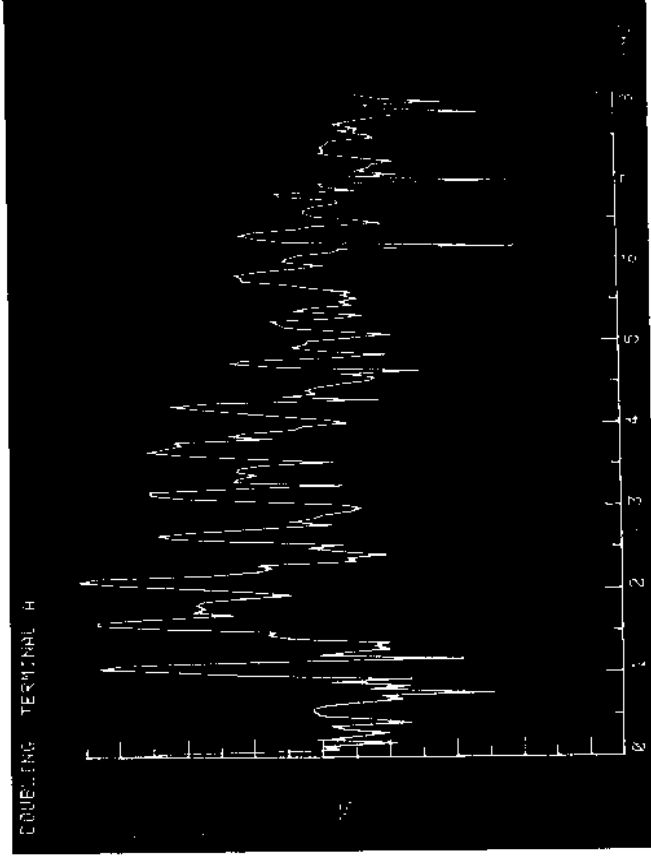


Figure 2f. DFT centered 5 msec after third arrow in 2a (263 msec).



### VOCAL REGISTERS

During the first year, children use a wide range of  $f_0$  values and at least three vocal registers in their prelinguistic speech. It is tempting to hypothesize that, in some sense, the children are experimenting with their larynxes to determine both the range of possible outputs and the linguistic meaning that can be attached to them.

Before 20 weeks, the children's vocalizations are mostly characterized by laryngeal maneuvers. The vegetative noises, such as coughing and burping, involve closing the glottis. The nonvegetative noises that I have called "speech" are similar in that there is usually an initial glottal stop [ʔ], followed by an indeterminate vowel, followed by another glottal stop or [h]. The vowels most used [ae, e, a, U] are "indeterminate" because they cannot be uniquely specified by a particular tongue position. They involve minimal changes from the rest position and the proper formant frequencies can occur as the result of lip and larynx movements (Lieberman 1977). Which vowel actually is heard in a given utterance may be quite accidental, simply a function of whatever position these articulators happen to be in. The large proportion of fry register and breathiness in these utterances would be the result of only the interarytenoid muscles contracting. This type of contraction closes only the rear part of the glottis, so the vibration of the folds is slow and with very little tension (van den Berg 1968a). This is the vibratory pattern of fry register -- slow and irregular pitch pulses.

Fry register can have two quite different linguistic uses. One is as a segmental feature of either vowels or consonants. Ladefoged (1971) says that Chadic languages, such as Housa, Bura, and Margii, have "laryngealized" stops and semivowels, while Nilotic languages, like Ateso and Lango, have laryngealized vowel phonemes. Laryngealization is Ladefoged's term for the use of fry (creaky voice) in a segment. The other use is prosodic; fry register can be used at the end of a breath group to effect the  $f_0$  drop required in a "falling" intonation contour. The use of fry in this way, of course, enhances the pitch drop. Atkinson (1973) derived  $f_0$  contours for a number of speakers reading different sentence types. In several cases, the  $f_0$  fell to a fry register value at the end of

declarative sentences. In addition, our tape recordings reveal varying, but substantial, use of fry by parents, especially at the end of a breath group. However, fry register seems to be used intermittently throughout discourse by a speaker to produce an effect of overall lower  $f_0$ , for example, by women or by men with naturally high  $f_0$ 's. The study of the social conditions under which this happens would be an interesting one.

Leaving aside this last suggestion, we can consider our data in light of adult use of fry phonemically and prosodically. It may be the case that all infants babble in fry register, but that for some infants this fry register use serves as preparation for linguistic manipulation later on. It would be interesting to gather data on the use of fry by children learning languages such as those cited above to see how similar the early babbling is, as well as how the phonemic use of fry comes in later. In this regard, the use of fry over very brief portions of an utterance by our infants is relevant, because this kind of fry may be channeled into a segmental distinction more easily than utterance-long fry. Prosodically, children's fry often occurs as they begin and end modal phonation as if they are unable to do so smoothly. Observation of older children speaking in sentences suggests that both fry register and sudden voicelessness (even of vowels) are used to ensure the correct  $f_0$  contour. These phenomena may, in some sense, be "mistakes" on the children's part, yet they produce the desired effect. Further work on speech acquisition should include some study on how fry register comes to be used within an intonation system.

The production of different vocal registers is discussed by van den Berg (1968 a,b) and Hollien (1974) among others. Van den Berg is concerned primarily with the particular muscles that contract for each register; Hollien has focused more on vocal fold length and thickness. Van den Berg's description of "falsetto" register in adults is especially relevant to our data. In this register the vocal ligaments are maximally tense as a result of medial compression and contraction of the interarytenoid and cricothyroid muscles. Only the edges, not the body, of the vocal folds vibrate and the glottis does not close completely. The resulting signal should be almost a pure sine wave.

Figure 3a shows the narrow band spectrogram of an utterance with peak  $f_0$  of about 1000 Hz. Figure 3b

de range of  
s in their  
ypotheseize  
erimenting  
ne range of  
that can be  
ations are  
vers. The  
d burping,  
tive noises  
that there  
lowed by an  
lotal stop  
, A, U] are  
: uniquely  
ion. They  
ion and the  
e result of  
Which vowel  
ay be quite  
er position  
The large  
as in these  
only the  
is type of  
he glottis,  
l with very  
his is the  
d irregular

: linguistic  
cher vowels  
hat Chadic  
rgii, have  
ile Nilotic  
:yngealized  
foged's term  
gment. The  
used at the  
required in  
fry in this  
Atkinson  
of speakers  
eral cases,  
the end of

Figure 3a. Narrow-band spectrogram of utterance by JD at 33 weeks of [e:heh], with peak FF of about 1000 Hz.

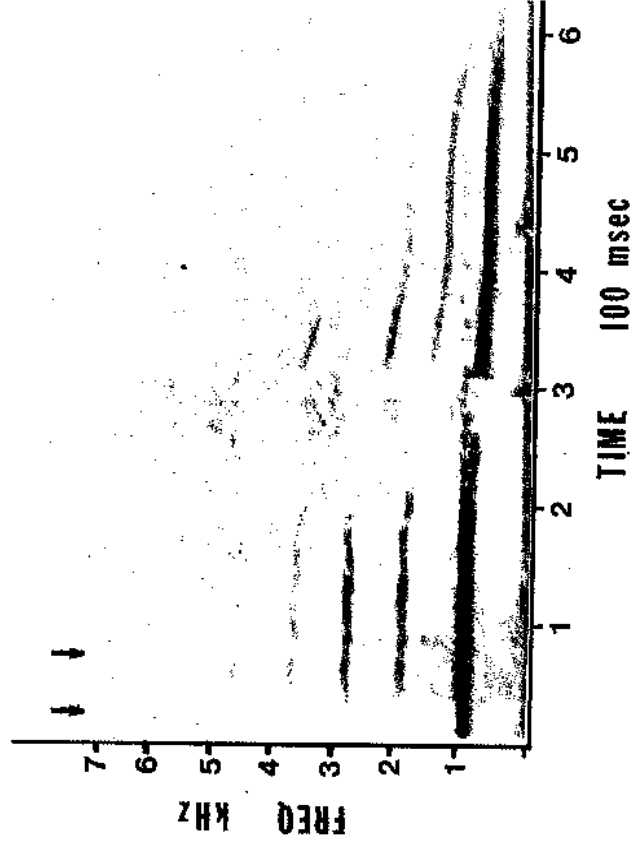


Figure 3b. DFT centered at first arrow in 3a (25 msec).

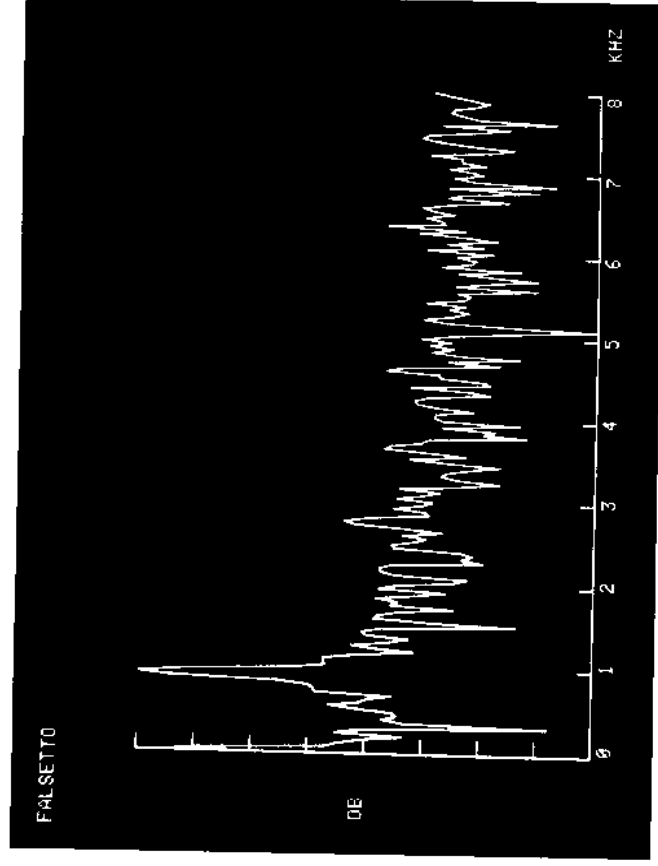


Figure 3c. Waveform corresponding to 3b.

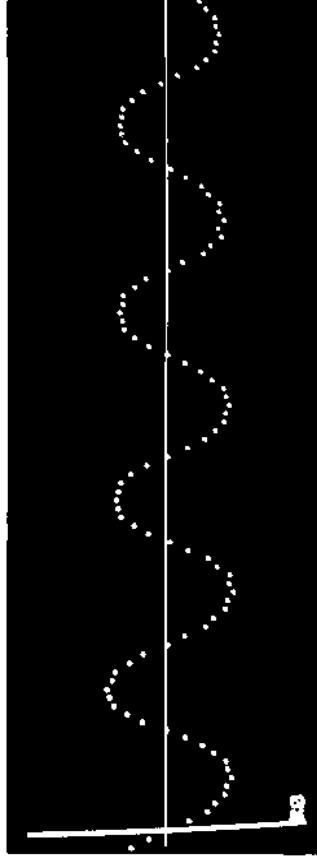
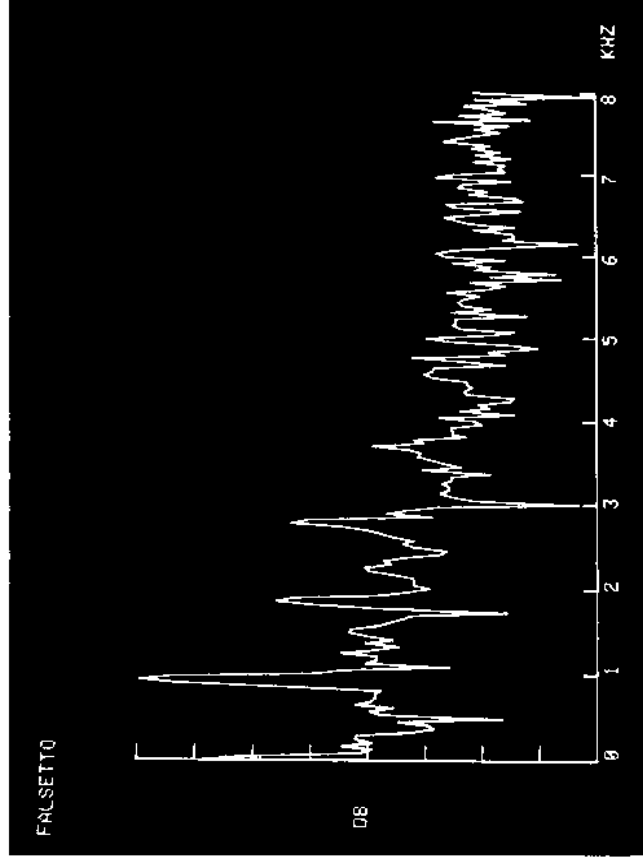


Figure 3d. DFT centered at second arrow in 3a (72 msec).



shows the Fourier Transform at the point marked by the arrow in 3a. This display shows that, at the beginning of the utterance, the signal is indeed almost a pure sine wave since there is a fundamental but no harmonics. The sinusoidal form is also apparent in Figure 3c, the waveform of the same portion of the utterance. The computer display has been stretched out for this picture; each point still represents .05 msec. Figure 3d shows the Fourier Transform of the portion centered at 72 msec in 3a. By this point in the utterance, three harmonics have been added in the glottal source. Note the large amplitude decrease across these harmonics; the source spectrum normally falls off more rapidly in high register than in modal.

The utterance with a peak  $f_0$  of 2500 Hz shown in Figure 4a is quite different in these respects. The waveform shown in Figure 4b is not sinusoidal. The pitch period of .4 msec corresponds to a  $f_0$  of 2,500 Hz. This  $f_0$  is also clear in the Fourier Transform shown in Figure 4c. Note that the harmonics do not fall off as they do in Figure 3d. These differences may be further evidence that a different register is involved in such utterances.

Van den Berg describes a "whistle" or "flute" register, in which the cricoarytenoid muscles supply medial compression on the vocal folds. This causes the glottis to close but leaves a small opening between the arytenoids. The vocal folds do not vibrate in this register; rather, air vibrates in the small cavity and so the  $f_0$  generated depends on the resonance frequency of this cavity. In van den Berg's excised adult larynxes, this  $f_0$  went as high as 2500 Hz. This mechanism may be responsible for the very high  $f_0$  values in our data. Interestingly, van den Berg offers

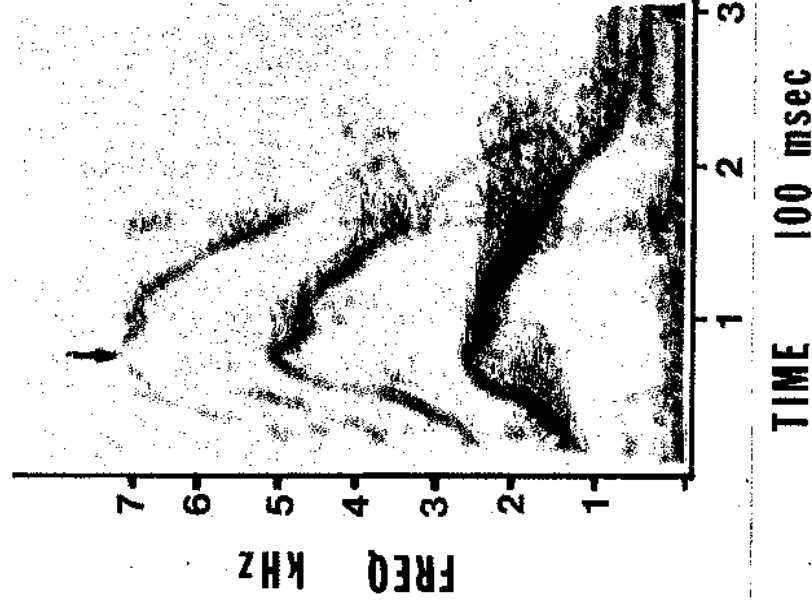


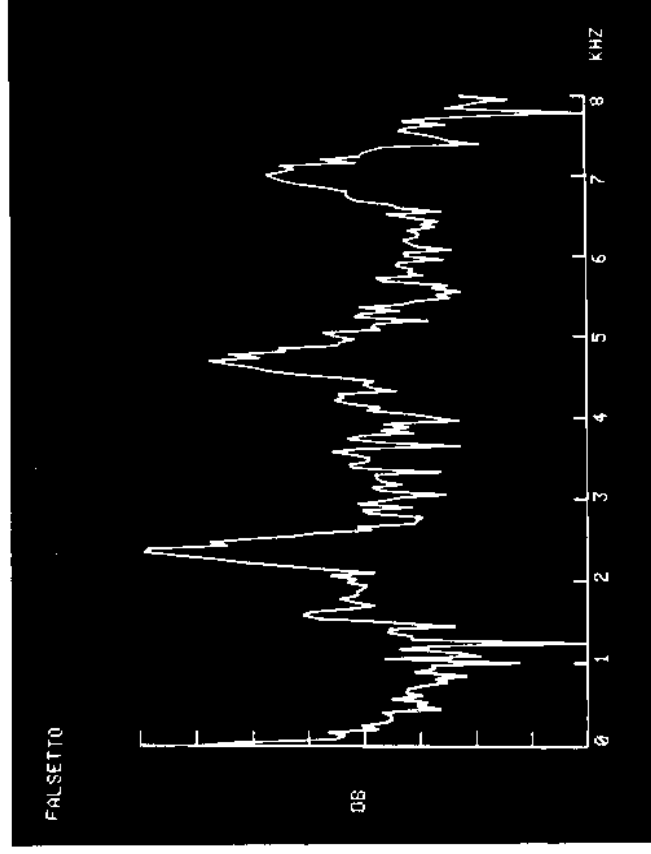
Figure 4a. Narrow-band spectrogram of utterance by JD at 33 weeks of [ae ʌ e], with peak  $f_0$  of about 2,500 Hz.

Hz shown in  
 aspects. The  
 soidal. The  
 $f_0$  of 2,500  
 er Transform  
 onics do not  
 differences  
 register is  
 ' or "flute"  
 scales supply  
 is causes the  
 g between the  
 rate in this  
 ll cavity and  
 nce frequency  
 xcised adult  
 00 Hz. This  
 very high  $f_0$   
 n Berg offers

Figure 4b. Waveform from arrow in 4a (70 msec).



Figure 4c. DFT corresponding to 4b.



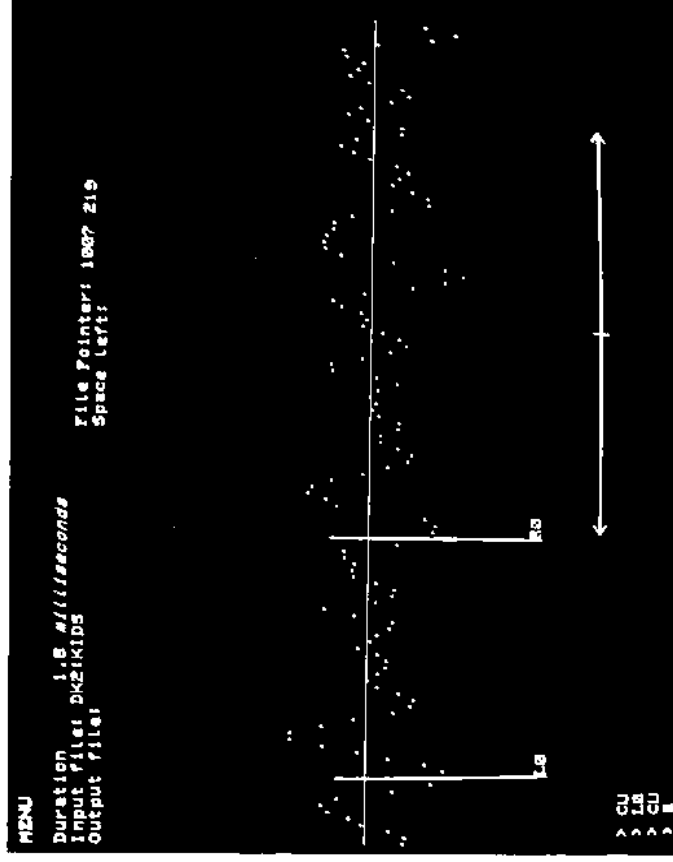
a way of determining-- in theory-- when this mechanism is used. Since the  $f_0$  is a resonance frequency, it would vary with the sound velocity of the gas medium used. The frequency of vocal fold vibration is not affected by breathing helium, for example, while formant (resonance) frequencies are doubled. If the  $f_0$  were to reach values such as 5000 Hz while an infant was breathing helium, then we could guess that it was not generated by vocal fold vibration.

A number of utterances involving what I have called doubling were analyzed with the Discrete Fourier Transform (DFT) analysis implemented on the Phonetic Laboratory's PDP-11/34 computer. The DFT displays corresponding to various points in time in an utterance usually show a consistent pattern: first a clear  $f_0$  at a high value, later a few relatively low-intensity harmonic peaks emerging from between the harmonics

(especially between 1 and 3 kHz), then clear harmonics of a  $f_0$  half that of the original one, except that the fundamental itself is missing, finally optionally a clear  $f_0$  at the halved value. In the intermediate stages there are invariably some displays for which it is difficult to determine the  $f_0$ . Measuring the pitch period from the waveform will usually put the  $f_0$  at the lower value in these cases. However, even the waveform is sometimes ambiguous in this regard.

Under these circumstances, it is difficult to say in what way and at what moment the  $f_0$  has changed. There is no auditory impression of a pitch drop, but it may be that it is so brief that it is not perceived. Since so many instances of doubling are associated with register shifts, it would be strange if the  $f_0$  went from, for example, a high modal value to half that value and then to a high register value. Still, in some parts of most doubling utterances one would measure the  $f_0$  as being halved. It is primarily the fact that the doubling harmonics do not continue throughout the utterance but, rather, fade in and out that led us to consider doubling a separate phenomenon. Even when the lower fundamental is present, the even numbered harmonics assert themselves in the pitch

Figure 5. Example of waveform with main and secondary pitch periods.



ar harmonics  
apt that the  
optionally a  
ntermediate  
for which it  
ng the pitch  
he  $f_0$  at the  
the waveform

lt to say in  
nged. There  
, but it may  
ived. Since  
ciated with  
the  $f_0$  went  
o half that  
Still, in  
; one would  
imilarly the  
ot continue  
le in and out  
a phenomenon.  
nt, the even  
n the pitch  
secondary pitch

period. Figure 5 shows a typical waveform display of this sort. The duration from peak to peak is 1.8 msec, corresponding to a  $f_0$  of 550 Hz, but there is also a secondary peak in this period. The duration from main peak to secondary peak is exactly half of that, .9 msec.

Possible mechanisms to account for doubling include sympathetic vibrations of some structure (e.g. the false vocal cords) with the vocal folds at certain unstable frequency regions. Alternatively, van den Berg (1968a, p. 24) noted that "thick and firmly adducted" vocal folds were sometimes observed to vibrate such that the body of the folds had a frequency twice that of the glottis. It is possible that such a phenomenon is involved here. However, the most likely mechanism is an alternating pattern of vocal fold vibration described for deaf speech by Monsen et al. (1979). For normal children, the question remains, whether certain phonetic contexts, utterance positions, or similar factors contribute to the occurrence of doubling.

This chapter is intended as a suggestion for future work to be done in the area of register use in speech. Both the study of register use in adult speech and its development in children's speech are inviting areas that deserve further investigation.

#### ACKNOWLEDGMENTS

This research was carried out while the author was a graduate research assistant in the Linguistics Department at Brown University and was supported by National Institute of Child Health and Human Development grant #5R01HD09197 to Brown University. I wish to thank Philip Lieberman for his advice and assistance, Robert Buhr for his part collaborations with me in this project; Cathy Kubaska and Howard Golub for helpful discussions; John Mertus for computer programs used in the analysis; and the mothers and children who volunteered for the study.



## REFERENCES

- Atkinson, J. 1973. Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. Ph.D. Dissertation. Storrs: University of Connecticut.
- ✓ Buhr, R. and Keating, P. 1977. Spectrographic effects of register shifts in speech production. J. Acoust. Soc. Am. 62:S25(A).
- Fairbanks, G. 1942. An acoustical study of the pitch of infant hunger wails. Child Dev. 13:227-32.
- Fairbanks, G., Herbert, E.L., and Hammond, J.M. 1949. An acoustical study of vocal pitch in seven- and eight-year-old girls. Child Dev. 20:71-79.
- Fairbanks, G, Wiley, J.H., and Lassman, F.M. 1949. An acoustical study of vocal pitch in seven- and eight-year-old boys. Child Dev. 20:62-69.
- ✓ Hollien, H. 1974. On vocal registers. J. Phonetics 2:124-43.
- Jones, M.C. 1965. An investigation of certain acoustic parameters of the crying vocalizations of young deaf children. Ph.D. Dissertation. Ann Arbor, Michigan: University microfilms.
- ✓ Keating, P. and Buhr, R. 1978. Fundamental frequency in the speech of infants and children. J. Acoust. Soc. Am. 63:567-71.
- Ladefoged, P. 1971. Preliminaries to Linguistic Phonetics. Chicago: University of Chicago Press.
- Lieberman, P. 1977. Speech Physiology and Acoustic Phonetics: An Introduction. New York: Macmillan & Co.
- Monsen, R.B. 1979. Acoustic qualities of phonation in young hearing-impaired children. J. Speech Hear. Res. 22:270-88.
- ✓ Monsen, R.B., Engebretson, A.M., and Vemula, N.R. 1979. Some effects of deafness on the generation of voice. J. Acoust. Soc. Am. 66:1680-90.
- Murry, T., Amundson, P., and Hollien, H. 1977. Acoustical characteristics of infant cries: fundamental frequency. J. Child Lang. 4:321-28.
- Ostwald, R.F., Phibbs, R., and Fox, S. 1968. Diagnostic use of infant cry. Biol. Neonat. 13:68-82.
- Peterson, G.E. and Barney, H.L. 1952. Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24:175-84.
- Prescott, R. 1975. Infant cry sound: developmental features. J. Acoust. Soc. Am. 57:1186-91.

1 in Speech:  
1 Study of  
on. Storrs:

phic effects  
J. Acoust.

of the pitch  
27-32.

i, J.M. 1949.  
seven- and  
79.

M. 1949. An  
- and eight-

J. Phonetics

of certain  
lizations of  
Ann Arbor,

al frequency  
J. Acoust.

Linguistic  
ago Press.  
and Acoustic  
Macmillan &

phonation in  
Speech Hear.

Temula, N.R.  
eneration of

n, H. 1977.  
fant cries:  
:321-28.

r, S. 1968.  
onat. 13:68-

ontrol methods  
1st. Soc. Am.

velopmental  
il.

Sheppard, W.C. and Lane, H.L. 1968. Development of  
the prosodic features of infant vocalizing. J.  
Speech Hear. Res. 11:94-108.

Stark, R.E. and Nathanson, S. 1975. Unusual features  
of cry in an infant dying suddenly and unexpectedly.

In Development of Upper Respiratory Anatomy and  
Function: Implications for Sudden Infant Death  
Syndrome, eds. J.F. Bosma and J. Showacre.  
Washington, D.C.: U.S. GPO. p. 237.

Truby, H.M., and Lind, J. 1965. Cry sounds of the  
newborn infant. In Newborn Infant Cry, ed. J. Lind.  
Uppsala: Almqvist and Wiksell. Chapter 1.

van den Berg, J. 1968a. Sound production in isolated  
human larynges. Ann. N.Y. Acad. Sci. 155:18-27.

van den Berg, J. 1968b. Register Problems. Ann. N.Y.  
Acad. Sci. 155:129-34.