Intrinsic $F0$ of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants

D. H. Whalen
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

941

Andrea G. Levitt
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Department of French, Wellesley College, Wellesley, Massachusetts 02181

Pai-Ling Hsiao
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Department of Linguistics, University of Connecticut, Storrs, Connecticut 06269

Iris Smorodinsky
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Department of Linguistics, Yale University, New Haven, Connecticut 06510

(Received 31 March 1994; accepted for publication 28 November 1994)

In every language so far examined, high vowels such as [i] and [u] tend to have higher fundamental frequencies ($F0$s) than low vowels such as [a]. This intrinsic $F0$ effect ($IF0$) has been found in the speech of children at various stages of development, except in the one previous study of babbling. The present study is based on a larger set of utterances from more subjects (six French- and six English-learning infants), at the ages 6, 9, and 12 months. It is found, instead, that $IF0$ appears even in babbling. There is no indication in these data of a developmental trend for the effect, and no indication of a difference due to the target language. These results support the claim that $IF0$ is an automatic consequence of producing vowels.

PACS numbers: 43.70.Ep, 43.70.Fq

INTRODUCTION

The relationship between vowel height and fundamental frequency ($F0$) has been noted for at least 60 years (Taylor, 1933). High vowels such as [i] and [u] tend to have higher $F0$'s than low vowels such as [a] and [æ]. The mechanism for this "intrinsic $F0$" ($IF0$) or "intrinsic pitch" has been the subject of great dispute (see the reviews in Ohala and Eukel, 1987; Sapir, 1989; Fischer-Jørgensen, 1990), but the consistency of the effect is not in question (Whalen and Levitt, in press). Every language that has been examined for $IF0$ (31 are listed in that work) has been found to have it, and these languages represent 11 of the world's 29 major language families. $IF0$ has been found not only in languages such as English (Peterson and Barney, 1952) and French (Di-Cristo, 1982) that use $F0$ primarily for stress and intonation, but also in tone languages such as Mandarin (Shi and Zhang, 1987) that use $F0$ changes to distinguish words. $IF0$ seems to be insensitive to the size of the vowel inventory as well, since both small (e.g., Japanese with 5 vowels) and large (e.g., German with 14) systems show similar effects (Whalen and Levitt, in press).

With such universality, $IF0$ has typically been assumed to be an automatic consequence of vowel articulation. Indeed, the theories reviewed in Sapir (1989) take this as a given. Under that assumption, it is of great interest whether the vowels of babbling will show this effect, since the babbling child presumably has no vowel categories *per se*, but simply vocalic articulations. If the child's vocal apparatus already has the interconnections that produce the $IF0$ effect in adults, then we should see $IF0$ in babbling. If there are significant anatomical or coordinative differences between infants and adults, perhaps $IF0$ will not appear in babbling.

Only one study that we have found has examined this question (Bauer, 1988). Bauer examined three infants at 9 and 13 months. There were 201 vowels measured at 13 months and an unreported number for the earlier age. Vowels were put into one of four broad classifications: high front, high back, low front, or low back. Bauer found no effect of height, but did find an effect of front/back. He attributed this to the high position of the larynx in the infant (Crelin, 1987). This high position also leads to a more vertical orientation, which might lead to more influence of the tongue pulling in the front/back dimension.

As a note of caution, though, the number of subjects and the number of tokens in Bauer's study were both rather small. The size of the study can greatly affect the outcome, as can be seen in a similar failure to find a vowel height effect, this time in running speech. Umeda (1981) measured approximately 200 vowels from two speakers in spontaneous conversation. She found no evidence of $IF0$ and concluded that it was not present in running speech. However, there are a great many factors that influence $F0$ in speech, and these were not controlled for in her study. To counteract this variability, it is necessary either to increase the number of observations, or to control the context. When factors such as sentence focus and segmental environment are properly controlled, even running speech shows the effect (Ladd and Silverman, 1984; Shadle, 1985). We can presume, then, that a larger sample of unrestricted text would show the effect.

 0001-4966/95/97(4)/2533/7/$6.00

TABLE I. Average $F0$ values for high versus low vowels for five studies that include children. Adult values for English (Peterson and Barney, 1952) and French (DiCristo, 1982) are given for comparison. Age is in years. $F0$ is in hertz (Hz). The difference is given both as a Hz value and (in parentheses) as a percentage of the a/æ value.

Study	Age (yrs)	N	Sex	$F0$ for i/u	$F0$ for a/æ	Difference
DiCristo, 1982	adult	1	fem	239	226	13(5.8)
	adult	3	male	133	124	9(7.3)
Peterson and Barney, 1952	adult	33	male	139	126	13(10.3)
	adult	28	fem	233	211	22(10.4)
	"child"	15	both	274	253	21(8.3)
Peterson, 1961	"child"	3	both	294	262	32(12.2)
Glaze *et al.*, 1990	5–11	97	both	250	229	21(9.2)
Hillenbrand *et al.*, in press	10–12	46	both	248	229	19(8.3)
Sorenson, 1989	6	3	male	290	258	32(12.4)
	6	3	fem	324	301	23(7.6)
	7	3	male	307	288	19(6.6)
	7	3	fem	288	279	9(3.2)
	8	3	male	267	255	12(4.7)
	8	3	fem	286	264	22(8.3)
	9	3	male	272	229	43(18.8)
	9	3	fem	300	275	25(9.1)
	10	3	male	263	243	20(8.2)
	10	3	fem	273	279	−6(−2.1)
(average)	6–10	30	both	287	267	20(7.5)

And, of course, it is not possible with babbling to restrict the context, so an increase in sample size is our only alternative. Thus the issue of $IF0$ in babbling cannot be considered to be settled, and the present study attempts to increase our understanding of this issue.

Although most researchers assume that $IF0$ is an automatic consequence of vowel production, others hold that $IF0$ is a deliberate enhancement of the speech signal by the speaker (Diehl and Kluender, 1989; Diehl, 1991). This account assumes that the perception of vowel height is a function not only of $F1$ frequency but of the difference between $F0$ and $F1$ (Traunmüller, 1981) and that speakers intentionally increase their $F0$ with high vowels to make this difference larger that it would otherwise have been. The universality of $IF0$, on this account, only argues for the usefulness of this particular enhancement. In babbling, however, there is no communicative intent and thus no distinctions to enhance. So the enhancement account should predict that $IF0$ will not appear in babbling. Even Bauer's (1988) finding, if it is correct, would be inconsistent with the enhancement account, since it implies an automatic (though different) mechanism for $IF0$. If $IF0$ were not found for babbling, the enhancement account would seem to be supported, with the assumption that $IF0$ would be an enhancement acquired later in development.

If $IF0$ is found for babbling, the most likely explanation is that it is not only universal but automatic. For the enhancement account to accommodate such a result, it would seem that an imitative explanation would be necessary. That is, since children hear this vowel height/$F0$ correlation in whatever adult language they hear, they include it in their babbling. Enhancement *per se* should not be an issue, since there are (presumably) no categories to enhance, but the imitation might be complex enough to include small $F0$ changes. This issue will be addressed further in Sec. III.

If enhancement is operative, we might expect there to be a developmental trend toward increased usage of the enhancement. Previous studies of $IF0$ in older children, with ages ranging from 5 to 11 years, show no indication of a developmental trend in $IF0$. Table I presents results from five published studies (Peterson and Barney, 1952; Peterson, 1961; Sorenson, 1989; Glaze *et al.*, 1990; Hillenbrand *et al.*, in press). We have averaged the two high vowels [i] and [u] and the two low vowels [a] and [æ]. As can be seen in the difference column of Table I, there is variability, especially in the Sorenson values where the N was small (only 3 per cell). But there is no indication of an overall trend toward larger (or smaller) effects.

If $IF0$ is universal, then we would expect to find similar patterns in the babbling of infants from any language environment. If $IF0$ is deliberate enhancement, we might expect that different languages would use the enhancement to different degrees. This difference might then appear as a difference in the babbling behavior of children in different language communities. The present study takes a first step in assessing the universality of $IF0$ in babbling by examining infants in two language environments, English and French. We have already found intonational differences between these two language groups in an earlier study (Whalen *et al.*, 1991). That study included 10 of the 12 subjects analyzed here. Since these children are using $F0$ in different ways in their babbling, it is certainly possible that they would treat $IF0$ differently if they were producing $IF0$ deliberately.

$IF0$ in babbling, then, needs further examination. The present study examines the babbling of 12 infants, 6 each in English and French environments. The infants were recorded in the home at 6, 9, and 12 months of age. We measured all the vowels except for the central (e.g., [ə]) and lower-mid (e.g., [ɛ] and [ɔ]). This resulted in 7325 tokens to analyze. With a larger set of results, we can more confidently address

TABLE II. Number of tokens analyzed for the 12 subjects. Speaker JZ had no recordings at 9 and 12 months, and MB had none at 12 months.

Language	French						English					
Initials Sex	MS M	NM F	YC F	JZ M	EC M	MB F	MM M	VB F	MA M	AB F	CR M	NG F
No. at 6 mo.	197	10	161	278	72	95	28	293	61	1250	80	130
No. at 9 mo.	105	33	353	···	37	83	38	123	30	708	130	337
No. at 12 mo.	147	29	850	···	171	···	48	205	226	647	136	234

the issue of whether *IF*0 is automatic or under the speaker's control.

I. METHODS

A. Subjects

The speakers were 12 infants, 6 learning French as their native language and 6 learning American English. The French infants were all living in Paris or its environs. The American infants lived in various cities on the northeast coast of the United States.

B. Stimuli

The utterances for the present study were selected from recordings made at weekly intervals by the parents of the children. Each infant was recorded in the home on a cassette tape recorder (Panasonic RQ 3145 or Marantz PMD 430) using a high quality microphone (Realistic supercardioid 33992A). Individual recording sessions lasted approximately 10–20 min. The parents were asked to choose a time when the child was likely to be alert and unlikely to cry. As far as possible, the microphone was held 20 cm from the baby. If necessary, the parent could attempt to induce babbling by speaking to the child (stopping, of course, when the infant began vocalizing). Additional comments about the session were recorded by the parent on a form provided with each tape.

All utterances from the 6-, 9-, and 12-month tapes were digitized onto the Haskins Laboratories VAX computer system. They were low-pass filtered at 9.6 kHz and sampled at 20 kHz, with preemphasis (Whalen *et al.*, 1990). We excluded cries, whispers, and various vegetative sounds. If an utterance contained a combination of speech and nonspeech, we would try to transcribe the speech.

All the utterances were transcribed by the third author, a native speaker of Mandarin Chinese. He is phonetically trained, and has experience with a wide variety of languages. Transcriptions were made from the digitized waveform, with the help either of the Haskins Laboratories program HADES (Rubin, 1995) or SIGNALYZE® (Keller, 1990). With these programs, the whole utterance could be heard repeatedly, as could any selected portion of the utterance. The overall character of the waveform also gave indications of possible syllables. The symbols of the International Phonetic Alphabet were used, with the understanding that some of the utterances would be very difficult to transcribe. We felt that obtaining a more detailed transcription was worth the effort involved, since this allows us to make more comparisons than Bauer's (1988) four-way classification.

Once the transcriptions were made, we selected the following vowels for analysis: high front ([i y ɪ ʏ]), midfront ([e ø]), low front ([æ a]), high back ([u ɯ ʊ]), midback ([o ɤ]), and low back ([ɑ ɒ ɐ]). (Strictly speaking, [ɐ is low central, but there were few enough members of this group anyway, so we included it.) While we did want to have the most accurate transcription possible, it was not possible to analyze the results any more finely than this, primarily because of the small number of instances of many of the vowels. There were a handful of tokens that were nasalized; these were simply included without any indication of the nasalization. We also treated all vowels without regard to their consonantal environment.

C. Analysis

All fundamental frequencies were measured from the speech waveform, by hand, using either HADES or SIGNALYZE®. The following procedure was used: For each syllable containing one of the vowels of interest, the main period of vowel activity was delimited. Then a location 40% of the way into this segment was found. In the best case, we would then measure five pitch periods to the left of the point and five to the right. The duration of this ten pitch period segment was then translated into an average *F*0 for that measurement point. In some cases, the pitch periods immediately around the 40% point were not measurable, either because the waveform was noisy or low in amplitude, or otherwise unclear. In those cases, the nearest ten measurable pitch periods within that syllable were chosen. Some tokens that had been transcribed proved to be too noisy or too faint to measure. Table II presents the number of measured tokens for the 12 subjects at the three ages. Two of the subjects lacked recordings at some of the months: JZ was missing the 9- and 12-month recordings, and MB was missing the 12-month recording. Both were French subjects. Another French subject, YC, lacked 12-month recordings but had 11-month ones. The 11-month recordings were used for the 12-month data for her. Two other subjects had sparse data at 1 or 2 months, so these were supplemented with recordings from an adjacent month. For English subject MA, 25.8% of the 6-month data was from the 6-month recordings, while the remaining 74.2% came from the 7-month recordings. Also for this subject, 37.7% of the 12-month data was from the 12-month recording, while the remaining 62.3% was from the 11-month recordings. Finally, for French subject MB, 35.0% of the 6-month data came from the 6-month recordings, while the remaining 65.0% came from the 7-month recordings.

TABLE III. $F0$ values for the six vowel categories for the 12 subjects.

	Front			Back		
	Mean $F0$	N	% of total	Mean $F0$	N	% of total
High	405.7	519	7.1	381.8	302	4.1
Mid	364.6	5254	71.7	359.9	89	1.2
Low	332.2	808	11.0	330.6	353	4.8

All of the target vowels were measured, with an exception for one subject. American subject NG had a large corpus, but the vast majority of her vowels were [e]. At six months, [e] was approximately 45× as frequent as the next vowel. At 9 months, the ratio was around 17 to 1. By 12 months, [e] outnumbered its nearest rival by a mere factor of 10. In order both to keep the representations of the vowels relatively similar, and to cut down on the amount of work required for this subject, only selected [e]'s were analyzed for her. For each age, a number of [e]'s was counted out (45, 17, or 10, for the three ages). The utterance containing that [e] was analyzed for all its [e]'s. Thus if there was only one [e], then that would be the only one analyzed. If that utterance happened to have several [e]'s, all of them were analyzed. In this way, [e] was still the most frequent vowel, but only by an overall factor of 2.3.

The distribution of these vowels is similar to those found in previous studies. In one cross-language study (Boysson-Bardies *et al.*, 1989), back vowels were found to be relatively rare (6.6% of the utterances in French and English), though the low back vowels were the most common of those. In the present study, by contrast, the high back vowels accounted for a higher proportion of the back vowels than was the case for the other study. (The proportion of front vowels overall would be higher if we had not excluded many of NG's [e] vowels.) Our proportions are more in agreement with Buhr's (1980) one English-learning infant. The selection criteria used here were too different to allow a direct comparison with the de Boysson-Bardies *et al.* (1989) study, but the distribution of vowels analyzed here is at least qualitatively similar to that found in other studies.

$F0$s larger than 700 Hz were excluded from the analysis. These represented 4.3% of the 7651 tokens measured. Such extreme values, while common in babbling, distort the means for those cells with small N's. It is also possible that a different phonation type is involved in such high $F0$'s, which would be a second reason to exclude these values. The selection process resulted in 7325 tokens being measured for the 12 subjects.

II. RESULTS

Means for the six vowel types for the 12 subjects are given in Table III. Also given are the number of tokens that went into each value. Table IV gives the size of the $IF0$ effect, both for height and front/back. The front/back difference is given as the front vowel mean minus the back vowel mean. In this way, any difference that matches the results of Bauer (1988) will be positive in value, while contrary results will be negative. As can be seen, there is a positive difference for height for 10 of the 12 subjects. For front/back, only 5 of the 12 subjects match Bauer's results.

For an analysis of variance, we operated on the means for each of the six cells for the 12 subjects. An analysis that used each observation was attempted, but the enormously large degree of freedom for the error term meant that almost any difference, however trivial, appeared significant. Using the means also gives the subjects with fewer productions a stronger say in the analysis. Since the differences among speakers, not tokens, are of primary importance, this result is to be desired.

The analysis, then, included the grouping variable language (English or French), and two within factors, height and front/back (with 3 and 2 levels, respectively). Three of the subjects (MM, NM, and MB) have missing cells, due to the lack of any instances of the mid back vowel category. Rather than reject these subjects from the analysis, these cells were replaced with the means of the five other cells for these subjects. This is a conservative approach to data replacement, since it will tend to minimize differences that actually exist. Language was not a significant main effect [$F(1,10)<1$, n.s.], indicating that the babblers had roughly equivalent overall $F0$'s. Height was a significant factor [$F(2,20)=16.62$, $p<0.001$], while the interaction with language was not [$F(2,20)=2.09$, n.s.]. Front/back was also not a significant factor [$F(1,10)<1$, n.s.]; neither was the interaction with language [$F(1,10)<1$, n.s.]. The two-way interaction of height and front/back was significant [$F(2,20)=4.48$, $p<0.05$], but the three-way interaction with language was not [$F(1,10)=1.09$, n.s.].

For the analysis by age, it was necessary to restrict the number of cells. By the time we break the results down into the six categories and the three ages for the 12 subjects, 45 of the 216 cells are empty. Most of these are for the low back and mid back vowels. Therefore we analyzed the four other cells, as a single factor of vowel quality with four levels, so that only differences among the four cells can be tested, not the front/back and high/low dimensions. Eighteen of the 45 empty cells come from the two subjects who lacked certain

TABLE IV. Size of the high/low difference and the front/back difference in $F0$ for the 12 subjects. The first and third rows are in hertz and the middle row is the high/low difference expressed as a percentage of the low vowel $F0$.

	French						English					
Initials	MS	NM	YC	JZ	EC	MB	MM	VB	MA	AB	CR	NG
High–low	56.0	94.3	40.4	18.4	−9.3	−26.7	15.6	70.9	15.1	93.8	101.3	81.4
(as %)	16.8	23.9	12.6	5.1	−2.7	−6.4	3.7	21.0	4.2	27.0	35.8	26.0
Front–back	−11.1	−19.5	−16.7	−13.9	19.3	−35.7	20.1	−28.7	−15.4	34.3	7.0	54.9

TABLE V. Mean $F0$'s for four of the six vowel categories for the ten subjects that had measurements at each of the months analyzed. The midback and low back categories were missing for many of the subjects for one month or another. The last row shows the difference between the mean of the two high vowel categories and the low vowels category.

Age in months	6		9		12	
	Mean $F0$	N	Mean $F0$	N	Mean $F0$	N
High back	373.5	88	369.8	71	394.6	125
High front	402.3	115	412.5	171	403.0	221
Midfront	350.4	1757	381.8	1311	364.1	1838
Lowfront	313.6	227	336.7	190	336.9	327
Diff. between Highs and low	74.3		54.5		61.9	

months, as mentioned before: JZ had no 9- or 12-month data, and MB had no 12-month data. These two subjects were excluded from this analysis, so that the remaining subjects had no missing cells. The ANOVA factors were vowel, with four levels, language, with two levels (English and French), and age, with three levels (6, 9, and 12 months).

In this analysis, as before, language was not a significant factor [$F(1,8)<1$, n.s.]. It did not enter into any significant interactions either. Age was not a significant main effect [$F(2,16)<1$, n.s.], which is to be expected: Even though $F0$ lowers throughout development (see Table I), the time elapsed here is too short to show this effect. Vowel is a significant main effect [$F(3,24)=8.22$, $p<0.001$], again showing the height effect. The critical interaction, age by vowel, is not significant [$F(6,48)=1.97$, n.s.], giving no indication of a difference in the effect over the six months involved here (see Table V). Even if we analyze each month separately (despite the lack of an interaction), the $IF0$ effect is present at each age. The separate analyses are strong for the 6-month [$F(3,27)=9.61$, $p<0.001$) and 12-month [$F(3,27)=5.05$, $p<0.01$] measurements, and somewhat less robust for the 9-month [$F(3,27)=2.94$, $p=0.0512$]. There is no evidence of change in $IF0$ over this time span.

Since we relied on our transcriptions to separate the vowels into categories, we need to be sure that we can do this independently of $F0$. In adult speech, it is certainly clear that different vowels can be produced with a wide range of $F0$'s without losing the vowel's identity. With babbling, however, it is not possible to ask the speaker to reproduce a particular vowel. One way of avoiding the vowel identity problem would be to correlate $F0$ with $F1$. Since $F1$ is lower with the high vowels and higher with the low vowels, there should be a negative correlation between $F0$ and $F1$ when $IF0$ is present. In the Peterson and Barney (1952) data, in fact, there is such a correlation if we examine the three speaker groups (the 33 adult males, the 28 adult females, and the 15 children) separately. When we correlate each individual production (there were two per vowel) for each vowel for all the speakers, we obtain the following correlations: males, $r=-0.16$ ($p<0.001$), females, $r=-0.20$ ($p<0.001$), and children, $r=-0.10$ ($p<0.10$). The correlation does not reach significance for the children either because of greater variability of their values or the smaller number of subjects.

When we examined our babbling data, however, there was a positive correlation between $F0$ and $F1$, but this was due to the fact that the formants were almost invariably excited by a single harmonic. With a mean $F0$ of 370 Hz, the formants in our set of babbles are poorly represented. If a harmonic happens to be at the center frequency of a formant, the two nearest harmonics would be approximately 15 db lower in amplitude even with a bandwidth of 100 Hz. [For adults, bandwidths typically remain in the range 50–60 Hz for formant values up to 2000 Hz (Dunn, 1961).] Harmonics with such low amplitudes are too close to the background level to contribute to the measurement of the formant. Occasionally in our measurements, we found two harmonics of equal amplitude, and it was possible to assume that the center frequency of the formant was between them. (With such limited measurements of the formant frequencies, it was, of course, impossible to measure the bandwidth with any confidence.) The appearance of two harmonics was uncommon, so the formant value was much more likely to be identical to one of the harmonics, resulting in the positive correlation between $F0$ and formant frequency.

As a final check on the possible misperception of $F0$ as vowel quality, we examined the distribution of the vowel categories by $F0$. If $F0$ were the only factor, then the distributions should be distinct. If any vowel can occur on any $F0$, then the distributions should be greatly overlapped. Figure 1 shows a highly overlapped pattern. For that figure, the number of tokens of a particular vowel in an $F0$ range or "bin" of approximately 16 Hz was counted. The top panel shows the proportion of all the vowels in the six categories. Because the midfront category is so disproportionately represented, it is hard to see the other distributions. Thus the lower panel shows the same data with a ceiling on the midfronts. As is clear, there are vowels of each category at every level of $F0$. Certainly the distributions are different, since that is what the $IF0$ effect consists of. But it is not the case that a high $F0$ was enough to cause a perception of a high vowel. The identifiability of the vowels was apparent throughout the $F0$ range. Thus there is no evidence of any large perceptual bias in the transcriptions.

It is impossible to rule out smaller perceptual biases which might have influenced the results. Indeed, small effects of $F0$ on the identification of ambiguous vowels have been found in one study by Reinholt Petersen (1986). Using synthetic vowels ranging from [u] to [o], he found that the most ambiguous vowel received more [u] responses with a high $F0$ compared with the low $F0$. The effects were quite small and never enough to change the majority decision. In addition, it was only possible to shift an ambiguous vowel from one category to a neighboring category. The results of Gottfried and Chew (1986), in which a wide range of $F0$'s for sung vowels was used, also show extremely few instances in which the height of the perceived vowel differs by more than one level. Thus even if there were bias effects in the present transcriptions, such biases would not account for the $F0$ difference between the low vowels and the high vowels. Despite the impossibility of completely ruling out small bias effects, then, the pattern of results strongly suggests that the effects we have found are due to the vowel articulation and not to the transcription.

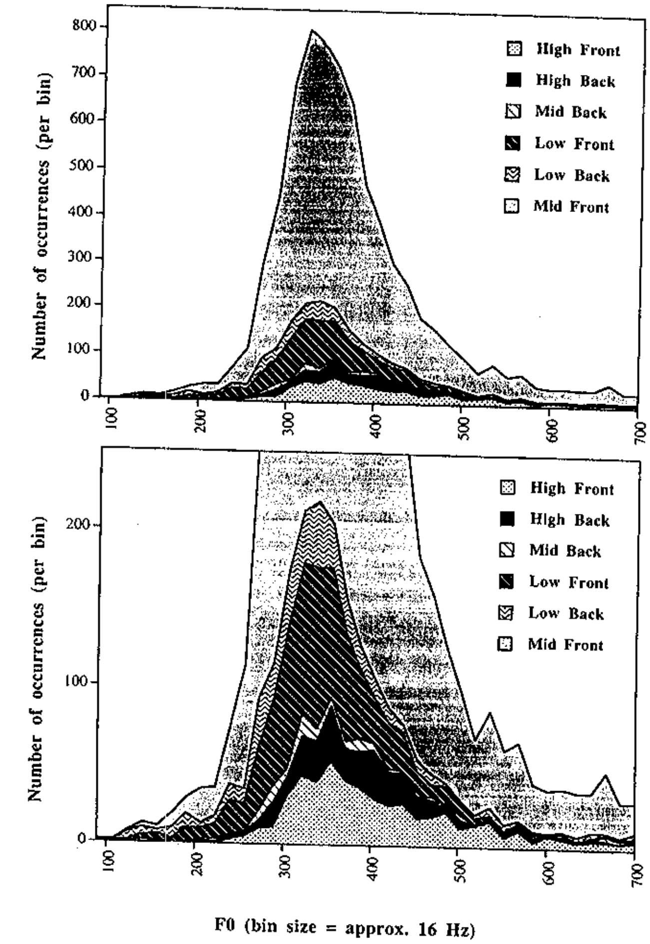

FIG. 1. Distribution of the six vowel categories by $F0$. Top panel: all vowels analyzed. Lower panel: the same data truncated at 250 occurrences, giving better resolution for the less well represented categories.

III. DISCUSSION

Our analysis of the babbling of 12 infants, 6 each from English- and French-learning situations, indicates that the intrinsic $F0$ ($IF0$) associated with vowels appears even in babbling. There was no evidence of a change in the effect across the three ages examined (6, 9, and 12 months). There was also no evidence of any difference between the two languages. These results are most compatible with the hypothesis that $IF0$ is an automatic consequence of vowel production.

The previous study that examined this question (Bauer, 1988) did not find a height difference, but instead found a front/back difference. That author attributed the fact that infants differ from adults to the relatively high position of the larynx in the vocal tract of young children (Crelin, 1987). However, we believe that his results differ from ours because of the scope of the studies. Bauer examined three children, and only 201 tokens at age 13 months. (He does not report the number of tokens for the 9- to 12-month portion.) This is too small a number to use for an unconstrained situation such as babbling. If we could have infants give us multiple repetitions with the same intonation, then a smaller number would be enough. But infants are constantly exploring the $F0$ range as they babble, and the placement of vowels of different qualities is random in this distribution. Thus it is very easy to have several utterances with high, even squealy pitch with a low vowel. It takes a large sample for this to average out. As can be seen in Table IV, there were two subjects who showed higher $F0$'s for low vowels, and they were two among those with the smallest number of tokens to analyze.

The differences between front and back vowels were not consistent from subject to subject in the present study. This is unlike Bauer's (1988) results but like the adult studies (Whalen and Levitt, in press). Given that Bauer's explanation of the front/back effect as due to the high larynx position is a plausible one, we need to explain why this high position does not change the $IF0$ effect. In fact, the high position seen in Crelin's x-ray images is somewhat misleading, since most of those were taken at rest. As Crelin himself notes (1987: p. 96), the larynx is pulled down into a much more adultlike position during speech (and screaming). That is, infants must work to make their vocal tracts appear more adultlike, and doing so seems to bring their larynx into the same relationship with the tongue that adults have. Since they show the same $IF0$ as adults, it seems likely that the same mechanism is involved as well. One might still suppose that different children might adopt different strategies, but even the two subjects who showed a contrary effect for height were inconsistent for front/back: Subject EC had the difference that Bauer found, while MB went in the other direction. So, as with the adults, there is no consistent effect of the front/back dimension on $F0$.

The present study also found no evidence of a developmental change over the 6-month span examined. This is consistent with the universality of $IF0$ (Whalen and Levitt, in press) and with the lack of any evidence of a developmental change later in life (Table I). The overall percentage of difference found for the babblers was 13.9% (as calculated from Table III). This is slightly higher than that found for other studies (Table I), but a statistical artifact is probably the cause. If we had transformed the measured $F0$'s into a semitone scale before averaging, the effect of the very high $F0$'s would have been reduced, and the difference between high and low vowels would probably have been much more similar, too. Certainly, if there is any developmental trend, it is for less $IF0$ rather than more. This does not fit with the enhancement hypothesis.

The present results are at odds with the one previous study (Bauer, 1988) and call into question the explanation given there. The difference is most likely due to the difference in sample size (200+ tokens versus the present 7000+). In addition, his developmental change was based on a comparison of the size of two different F ratios, which is not a reliable method of comparing results across data sets. It is also risky to assume that the absence of a significant difference means that there is no effect. However, the measurements here are sufficiently strong to show the $IF0$ effect at each of the three ages analyzed, so at least we know that the effect is not absent at any of the ages. It seems likeliest that there is no developmental change in $IF0$.

These results also cast doubt on the description of $IF0$ as a deliberate enhancement of the speech signal (Diehl and

Kluender, 1989; Diehl, 1991), for three reasons. First, while infants may begin to perceive the vowel categories of their target language at an early age (Kuhl *et al.*, 1992), the evidence from vowel productions in babbling shows only a tendency toward the target language formant space (Boysson-Bardies *et al.*, 1989) or intonation (Whalen *et al.*, 1991), not toward specific categories. If the infants have no categories to enhance, why should they use $IF0$? It is true that infants will hear vowels with $IF0$, since every language shows $IF0$ (Whalen and Levitt, in press). So if they are imitating what they hear, then they might imitate $IF0$ differences. Infants do not imitate just the native vowel categories, though. They produce some non-native vowels and seldom if ever produce some of the native vowels. It is hard to see why they should imitate the $IF0$ feature, nor indeed how they might abstract away from the individual vowels to the more general principle that vowel height is what is important. Second, the $IF0$ effect disappears at the lower portion of adult speakers' ranges (Whalen and Levitt, in press). The explanation for this phenomenon is likely to come from the different mechanisms for raising and lowering $F0$. Unless the infants have already understood this difference, the lack of $IF0$ at low values would seem to add uncertainty to the generalization that imitative $IF0$ would have to be based on. Finally, the task of detecting $IF0$ in the course of running speech would seem especially difficult in the case of learners of a tone language, since they certainly hear a great deal of $F0$ variation that is important (the tones) that is completely unrelated to $IF0$. However, $IF0$ also occurs in tone languages like Mandarin (Shi and Zhang, 1987). We might expect that infants who are learning Mandarin as their native language would fail to use $IF0$ if any babblers would. Given the importance of tone in Mandarin and its independence from $IF0$, Mandarin-learning infants might not easily produce $IF0$ if it had to be learned. On the other hand, if $IF0$ is automatic, then even the Mandarin-learning infants would show $IF0$. This remains to be tested.

The present study shows $IF0$ in babbling. The effect is independent of which of the two target languages (French or English) were involved. There does not seem to be a developmental trend for the 6- to 12-month range examined. So, despite the fact that vowels can be produced with a wide range of $F0$s, it appears that $IF0$ is an automatic consequence of vowel articulation.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. DC-00403 to Haskins Laboratories and Catherine Best. Portions of this research were presented at the 126th Meeting of the Acoustical Society of America, Denver, Colorado, October 1993, and the 9th International Conference on Infant Studies, Paris, June 1994. Additional help with the stimuli was provided by Michele Sancier, Winifred McGowan, and Julia Irwin. We thank Arthur S. Abramson, Catherine T. Best, Hartmut Traunmüller, Keith Johnson, and an anonymous reviewer for helpful comments.

Bauer, H. R. (1988). "Vowel intrinsic pitch in infants," Folia Phoniat. **40**, 138–146.

Boysson-Bardies, B. de, Halle, P., Sagart, L., and Durand, C. (1989). "A crosslinguistic investigation of vowel formants in babbling," J. Child Lang. **16**, 1–17.

Buhr, R. D. (1980). "The emergence of vowels in an infant," J. Speech Hear. Res. **23**, 73–94.

Crelin, E. S. (1987). *The Human Vocal Tract: Anatomy, Function, Development, and Evolution* (Vantage, New York).

DiCristo, A. (1982). *Prolégomènes à l'étude de l'intonation: Micromélodie* (Editions du Centre National de la Recherche Scientifique, Paris).

Diehl, R. L. (1991). "The role of phonetics within the study of language," Phonetica **48**, 120–134.

Diehl, R. L., and Kluender, K. R. (1989). "On the objects of speech perception," Ecol. Psychol. **1**, 121–144.

Dunn, H. K. (1961). "Methods of measuring vowel formant bandwidths," J. Acoust. Soc. Am. **33**, 1737–1746.

Fischer-Jørgensen, E. (1990). "Intrinsic $F0$ in tense and lax vowels with special reference to German," Phonetica **47**, 99–140.

Glaze, L. E., Bless, D. M., and Susser, R. D. (1990). "Acoustic analysis of vowel and loudness differences in children's voice," J. Voice **4**, 37–44.

Gottfried, T. L., and Chew, S. L. (1986). "Intelligibility of vowels sung by a countertenor," J. Acoust. Soc. Am. **79**, 124–130.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (in press). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. (to be published).

Keller, E. (1990). SIGNALYZE, *Signal Analysis for Speech and Music: User's Manual* (InfoSignal, Inc., Rosemere, Quebec).

Kuhl, P. J., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.

Ladd, D. R., and Silverman, K. E. A. (1984). "Vowel intrinsic pitch in connected speech," Phonetica **41**, 31–40.

Ohala, J. J., and Eukel, B. W. (1987). "Explaining the intrinsic pitch of vowels," in *In Honor of Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris, Dordrecht), pp. 207–215.

Peterson, G. E. (1961). "Parameters of vowel quality," J. Speech Hear. Res. **4**, 10–29.

Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Reinholt Petersen, N. (1986). "Perceptual compensation for segmentally conditioned fundamental frequency perturbation," Phonetica **43**, 31–42.

Rubin, P. E. (1995). "HADES: a case study of the development of a signal analysis system," in *Behavioral Aspects of Speech Technology: Theory and Applications*, edited by R. Bennett, S. L. Greenspan, and A. Syrdal (Elsevier, Amsterdam), pp. 501–520.

Sapir, S. (1989). "The intrinsic pitch of vowels: Theoretical, physiological and clinical considerations," J. Voice **3**, 44–51.

Shadle, C. H. (1985). "Intrinsic fundamental frequency of vowels in sentence context," J. Acoust. Soc. Am. **78**, 1562–1567.

Shi, B., and Zhang, J. (1987). "Vowel intrinsic pitch in standard Chinese," in *Proceedings of the XIth ICPHS* (Academy of Sciences of the Estonian SSR, Tallinn), Vol. 1, pp. 142–145.

Sorenson, D. N. (1989). "A fundamental frequency investigation of children ages 6–10 years old," J. Commun. Disord. **22**, 115–123.

Taylor, H. C. (1933). "The fundamental pitch of English vowels," J. Exp. Psychol. **16**, 565–582.

Traunmüller, H. (1981). "Perceptual dimension of openness in vowels," J. Acoust. Soc. Am. **69**, 1465–1475.

Umeda, N. (1981). "Influence of segmental factors on fundamental frequency in fluent speech," J. Acoust. Soc. Am. **70**, 350–355.

Whalen, D. H., and Levitt, A. G. (in press). "The universality of intrinsic $F0$ of vowels," J. Phon.

Whalen, D. H., Levitt, A. G., and Wang, Q. (1991). "Intonational differences between the reduplicative babbling of French- and English-learning infants," J. Child Lang. **18**, 501–516.

Whalen, D. H., Wiley, E. R., Rubin, P. E., and Cooper, F. S. (1990). "The Haskins Laboratories' pulse code modulation (PCM) system," Behav. Res. Methods, Instrum. Comp. **22**, 550–559.