A RELIABILITY STUDY OF SPEECH SOUNDS OBSERVED IN THE CRYING OF NEWBORN INFANTS[1]

ORVIS C. IRWIN AND HAN PIAO CHEN

INTRODUCTION

In any research program the primary concern is to safeguard the data. The term "safeguarding the data" here is taken to mean that they are: 1) reliable and 2) valid. This study is concerned with the first of these two problems. Reliability usually is treated under two headings: 1) the problem of observer reliability and 2) the reliability of the data. The purpose of this study is to investigate both of these problems in regard to early speech sounds, especially when they occur during the crying of newborn infants.

The general status of the problem of the reliability of infant speech sound data has been reviewed by Irwin (1). A study of Irwin and Curry (2) reported on the reliability of recording vowel sounds of newborn infants, and Irwin and Krehbiel (3) in an unpublished study reported both observer and data reliability of transcriptions of vowel and consonants of four-, five-, and six-month infants. In all the studies now being conducted the International Phonetic Alphabet is used to transcribe sounds.

In the present study two experiments were conducted, one of which involved a small sample of crying sounds from each of 40 infants;[2] the other, a fairly large sample from each of 5 infants. The unit of observation was the respiration (2). In the first experiment the sample consisted of the sounds carried on each of 30 exhilations giving a total of 1200 breaths for the group. The total number of individual vowel elements on these 1200 respirations amounted to 1520. In the second experiment the sample consisted of the sounds carried on 180 breaths for each baby giving a total of 900 breaths for the group. The total number of separate vowel sounds amounted to 1080. In addition a further analysis of the Irwin-Curry data has been made. This sample consists of the sounds carried on 25 respirations of each of 40 infants or a total of 1000 breaths which yielded a total of 1285 vowel sounds. The data under analyses then consist of three samples with a grand total of 3885 vowel sounds.

WHINING AND CRYING

In the study by Irwin and Curry a discrimination was made between crying and non-crying sounds. However since non-crying sounds are of rare occurrence in newborns, no report can be made at present concerning them. In the two experiments of the present study a further discrimination was made between crying and whining sounds. A report on the latter will be deferred until later but it does involve the necessity or

[1] *From the Iowa Child Welfare Research Station, University of Iowa, Iowa City, Iowa.*

[2] *Ten of these subjects were negro infants. A systematic study of racial differences in speech sound equipment of newborns will be reported later.*

establishing observer reliability for discriminating these two categories. Accordingly this problem also was undertaken.

Two observers simultaneously but independently recorded 60 breaths on each of 5 infants. Speech sounds as such were disregarded and the record on each of the 300 breaths consisted either of the term "whining" or "crying." In determining the difference between whining and crying in newborns, the observers used the following set of criteria:

1. Regularity of respiration
2. Shape of mouth opening
3. Tongue position
4. Face and lid contraction
5. Loudness of sound

On the basis of these criteria the condition of crying is defined as regularity of breathing, the mouth is opened wide, usually in a rectangular shape, the tip of the tongue is elevated, the muscles of the face are strongly contracted and the eyelids tightly closed, and the sound is uttered with force or loudness. Whining is defined as occurring during irregular breathing, the mouth is partially opened, the tongue tip is not elevated, the muscles of the face are only slightly contracted, the lids are open or if closed are not tightly contracted, and the sounds are feeble.

The following tabulation presents the data on the 5 cases. That crying and whining may be reliably discriminated is evident. The per cent of agreement between the two observers ranged from 95 to 100, the average being 97 per cent.

Subject	Per Cent Agreement
1	97
2	95
3	100
4	97
5	95
Average	97

RELIABILITY IN TERMS OF FREQUENCY AND OF THE OBSERVED BEHAVIOR UNIT

The problem of observer reliability of speech sound data in newborns may be approached in two ways: 1) agreement in terms of the total number of speech elements considered as separate events, and 2) agreement in terms of the pattern of elements occurring on the respiration or breath. The former will be called frequency agreement, the latter breath agreement. In the case of the former the frequencies of each vowel element are calculated as a per cent of the total number of elements uttered regardless of patterning. This is done for each observer and the corresponding values for both are compared on a per cent basis.

A more rigorous analysis is concerned with the agreement between two observers' transcriptions when the pattern of sounds occurring on the breath is considered. The problem will be clarified by indicating several possible cases. The first case is presented when one observer hears the sound ʌ on a breath whereas the other records the sound ɛ.

This constitutes disagreement. In the second case one observer might hear the sounds ɛ, and I on a given breath and the second observer also might hear these sounds. Obviously this constitutes 100 per cent agreement. However there is a third possibility in which there is agreement on some of the sounds occurring on a breath and disagreement on others. One observer may record the sounds ɛ, ʌ on a given breath, whereas the other hears æ, ʌ. In this case, of a total of four sounds, two are disagreements and two are agreements. Accordingly the index is 50 per cent. This has been termed the mixed category.

All sounds, therefore, have been analyzed first into three categories: agreements, disagreements, and mixtures. Then the sounds classified in the mixed group in turn have been calculated in per cents of agreement and disagreement, and their values have been added to the original categories of agreement and disagreement. The three categories thus have been reduced to two and in this manner an index has been derived to indicate breath agreement.

EXPERIMENT 1

Table 1 gives the per cent of agreement for two observers for both breath and frequency in experiment 1 on each of 40 babies. The average agreement in terms of the breath unit is 91 per cent. For the frequencies it is 93 per cent.

The following tabulation indicates how the two sets of agreements are distributed. It will be seen that in the case of the breath unit there was agreement above 97 per cent on 12 infants, above 93 per cent on 18 infants, above 85 per cent on 33 of the 40 subjects. In terms of total frequency of sound elements there was agreement above 97 per cent on 17 infants, above 93 per cent with 27 infants, and above 85 per cent with 36 of the 40 subjects. In the case of breath agreement 6 indices

Agreement, Per Cent	Frequency	Cumulative Frequency
Breath Unit		
97 to 100	12	12
93 to 96	6	18
89 to 92	9	27
85 to 88	6	33
81 to 84	1	34
77 to 80	2	36
73 to 76	4	40
Total	40	
Total Frequency		
97 to 100	17	17
93 to 96	10	27
89 to 92	5	32
85 to 88	4	36
81 to 84	2	38
77 to 80	1	39
73 to 76	1	40
Total	40	

TABLE 1

Observer Reliability of Crying Sounds of Forty Babies

Infant Number	Frequency Agreement, Per Cent	Breath Agreement, Per Cent
1	74	76
2	98	74
3	87	77
4	100	100
5	94	74
6	94	86
7	80	88
8	86	87
9	89	79
10	82	89
11	98	96
12	100	93
13	100	100
14	82	76
15	93	89
16	92	82
17	100	100
18	100	98
19	86	85
20	94	98
21	97	99
22	95	91
23	97	99
24	97	90
25	94	96
26	95	88
27	100	96
28	97	97
29	100	100
30	87	90
31	95	89
32	92	88
33	90	91
34	89	92
35	93	91
36	100	94
37	97	94
38	100	100
39	98	97
40	95	97
Average	93	91

fall between 70 and 80 per cent whereas in the case of frequency agreement 2 indices fall between these limits.

Thus with either method a quite satisfactory observer reliability is established.

A graphic method of presenting the dispersion of speech elements involving the principle of the profile was described in a previous report (4). Figure 1 includes six profiles comparing the dispersion vowel sounds of newborns as recorded by three pairs of observers. The two upper profiles give the percent of frequency of observer I and observer

OBSERVED RELIABILITY ON PER CENT OF FREQUENCY OF VOWEL SOUNDS OF NEWBORNS

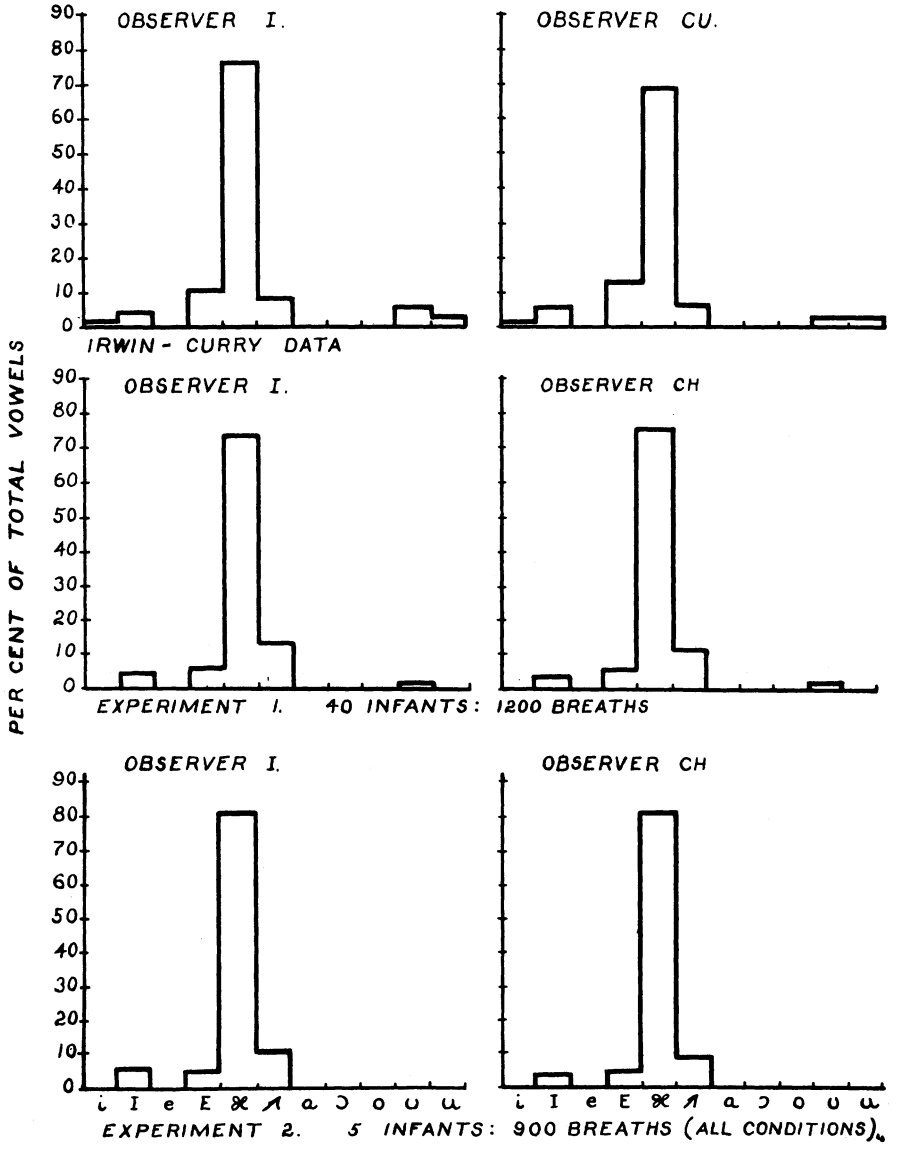

FIGURE 1

Cu. There is striking similarity between the patterns of these two observers. The two middle profiles were constructed from the data of experiment 1, and likewise show similarity of pattern. The two lower profiles are based on the data in experiment 2. Here again there is great similarity. The comparisons between the records of pairs of profiles quite adequately indicate the degree of observer reliability which may be achieved with this type of data.

It is of interest to note the trends which occur during the course of collecting the data. They are indicated in Figure 2. The curves represent the average agreements of eight successive groups of 5 babies. Thus the observer reliability on the first 5 babies is about 80 per cent. On the last 5 it is about 95 per cent. One curve represents the Irwin-Curry data, the other the Irwin-Chen data. While there are some differences between the two curves, they start together, they end together, and they indicate that an untrained observer will raise his reliability to over 90 per cent after working with about 18 infants. The figure thus is of practical value in conducting research on sounds for it tells approximately how many subjects should be used to train new observers.

There is a special problem involved in the derivation of the reliability index by the two methods. As was noted the per cent of breath agreement is 91, the per cent of frequency agreement is 93. Superficially it would seem that these two values are sufficiently close so that one method could be substituted for the other. That this is not the case will become apparent from the following reasoning.

When the two sets of values are correlated the coefficient of correlation is .75 $\pm$ .03. Ordinarily this might be considered a high correlation. However the coefficient of alienation, which measures the absence of relationship, amounts to .66. This value is too high to permit a substitution of one of these methods for the other. It may be inferred then that the two methods may not rank the infants alike. To test this inference, the index of forecasting efficiency may be applied. In order that one might predict scores derived by one of these methods from scores derived by the other method, a very high correlation is necessary. A correlation of .995 is required to reduce errors in prediction to 90 per cent. A correlation of 75 gives a percentage reduction over chance in errors of prediction of only 32.9. The two methods, therefore, do not rank the infants in the same order. This is another way of saying that the scores from the method of frequency agreement cannot be substituted for those derived by the method of breath agreement. A cautious interpretation of the correlation value of .75 therefore requires that both methods be used in computing observer reliabilities on speech data at the newborn level.

The reason that the two methods do not rank the individuals alike may lie in the fact that they represent two different problems. The method of breath agreement answers the question - what in terms of the observation unit (the single respiration) actually used is the observer agreement. The frequency method answers the question - what agreement may be achieved with a large amount of data collected over an extended period without regard for the observation unit. Quite evidently these are two different problems.

The Irwin-Curry data may be similarly interpreted. The correlation

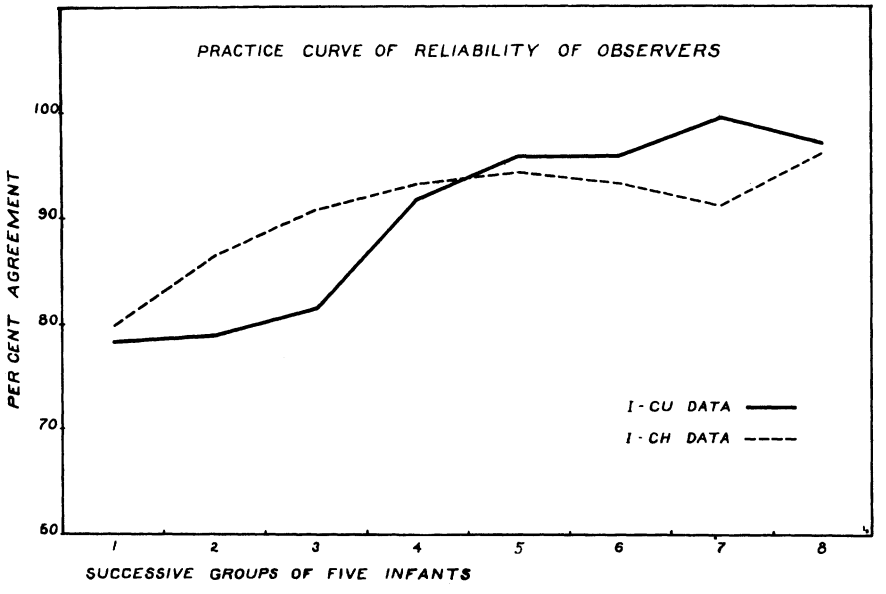

FIGURE 2

value between the breath and frequency ratings of this study amounts to .94. This is much higher than the previous value of .75. Nevertheless the index of forecasting efficiency for this correlation value is 65.8, not a very satisfactory value. The coefficient of alienation is .342 indicating the amount of absence of relationship. Here, too, evidence is forthcoming that the frequency method of establishing observer reliability cannot be substituted for breath reliability. The conclusion that both methods should be employed seems to be re-enforced by this study.

There is a fundamental criticism which may be leveled against this procedure of establishing observer reliability. It is that the untrained observer learns not only the correct habits of transcription from the trained observer but also he learns the errors of the trained observer. This means that the reliability index of 93 per cent is as high as it is because both correct and incorrect habits of transcribing are compared.

The situation has been safeguarded by requiring the trained observer (I) to take all courses in phonetics, both undergraduate and graduate, offered in the university. His original training was conducted by a trained phonetician. In addition he worked with several other experienced phoneticians so that his habits were not determined by a single individual. Moreover, in order to maintain a high degree of accuracy in the use of phonetic symbols, this observer has made it a policy to repeat attendance on these courses and engages in regular practice to maintain skill.

EXPERIMENT 2

In the second experiment 5 babies were used with a total of 180 observations on each. Thus it affords a comparison of a large number of observations on a small sample with the result of the first experiment in which 30 observations were made on each of 40 babies. In addition, this experiment was designed to answer an important question concerning the collection of speech data. In an earlier study (2) it was pointed out that the use of the short time sample failed to yield a satisfactory reliability. Accordingly a short behavior unit, the single breath carrying from one to five sounds, which falls well within the ordinary attention span of the observer, was used instead of a time sample. The individual respirations themselves, however, during violent crying occasionally may occur so rapidly that the observer cannot transcribe the sounds occurring on consecutive breaths. This is a fundamental criticism of the method of the short behavior sample. McCarthy (5) has made this criticism against the general method of phonetic transcription. That it can be circumvented, however, will be apparent from what follows.

A record of 60 breaths was taken simultaneously and independently under three different conditions by the two observers. Transcriptions were made of 60 consecutive breaths, of 60 alternate breaths, and of 60 fourth breaths. If a comparison of these three types of transcriptions should reveal striking differences then the criticism is a valid one. If, however, the per cents of agreement are well above 90 then every second or fourth, etc., breath, instead of consecutive breaths, may be transcribed with satisfactory results. That the latter is the case in terms of breath agreement is evident from the following tabulation which

Infant Number	Condition	Breath Agreement, Per Cent
1	Consecutive	85
	Alternative	84
	Every 4th	84
2	Consecutive	97
	Alternative	98
	Every 4th	97
3	Consecutive	96
	Alternative	96
	Every 4th	97
4	Consecutive	97
	Alternative	98
	Every 4th	98
5	Consecutive	97
	Alternative	98
	Every 4th	96
Average		95
Average	Consecutive	94
	Alternative	95
	Every 4th	94

presents observer reliability of crying sounds of the 5 babies under three conditions. All of the per cent values, it is evident, lie in the nineties except three on the first infant. It will be noted that the average agreements for the three conditions are practically the same.

In terms of frequency agreement the result is the same. This can be seen in the following tabulation.

Infant Number	Condition	Frequency Agreement, Per Cent
	Consecutive	87
1	Alternative	99
	Every 4th	86
	Consecutive	100
2	Alternative	97
	Every 4th	97
	Consecutive	96
3	Alternative	99
	Every 4th	99
	Consecutive	99
4	Alternative	95
	Every 4th	100
	Consecutive	97
5	Alternative	97
	Every 4th	96
Average		96

The average observer agreements under different conditions are:

Consecutive	96
Alternative	97
Every 4th	96

Moreover, so far as observer reliability is concerned, these tabulations reveal that a large sample of sounds from a small sample of subjects gives as satisfactory results as a small sample of sounds from a large sample of infants.

OBSERVER RELIABILITY OF THE FREQUENCY OF SEPARATE VOWELS

Another problem remains to be considered. What is the reliability of the observers on each of the vowel elements? The next tabulation considers this problem for the experiment which involved forty subjects. The second and third columns give the frequency of occurrence of each of five vowel sounds as recorded by the observers. The fourth column shows the per cent of agreement. It will be seen that there is very high agreement, 99 per cent, on the sound æ, which is the most frequent sound. The lowest per cent of agreement occurs in the sound I. Thus

the tabulation shows that for most of the sounds uttered by newborn infants there is good agreement. The average per cent of agreement is 85. When the frequencies themselves are transformed into per cents of the total number of sounds heard, the result is found in the last three columns.

Vowel	Frequency			Per Cent		
	Observer		Agreement, Per Cent	Observer		Agreement, Per Cent
	I	Ch		I	Ch	
I	79	47	60	5.1	3.2	63
ɛ	85	82	96	5.4	5.5	98
æ	1,165	1,153	99	74.7	78.0	96
ʌ	219	187	85	14.0	12.6	90
ʊ	12	10	83	.8	.7	88
Total	1,560	1,479		100.0	100.0	

The following tabulation presents similar analyses of reliability on separate vowels in the second experiment which involved five subjects. Only four sounds were heard during this experiment. The per cent of agreement varies from 82 on the vowel I to 99 on the sound æ. The average per cent agreement here is 89.

Vowel	Frequency			Per Cent		
	Observer		Agreement, Per Cent	Observer		Agreement, Per Cent
	I	Ch		I	Ch	
I	50	41	82	4.6	3.9	85
ɛ	46	52	88	4.2	4.9	86
æ	873	865	99	79.9	81.2	98
ʌ	124	107	86	11.3	10.0	89
Total	1,093	1,065		100.0	100.0	

DATA RELIABILITY

The problem of the reliability of the data will be handled in three ways: 1) by comparing sounds occurring on the consecutive, alternative and fourth breaths; 2) by comparing odd-even items; and 3) by comparing the results of three experiments.

Comparison of Consecutive, Alternative, and Fourth Respirations

In Figure 1 the lower pair of profiles was based upon data collected in experiment 2 under three conditions. These two profiles indicate the results of the three conditions when they are totalled. In Figure 3, however, a further analysis of these data is presented in profiles representing the dispersion of vowel sounds under each of the three conditions. The three profiles arranged vertically at the left show the patterns of the three samples taken consecutively, alternately and on every fourth breath by one of the observers. It is clear that with quite minor variations these three patterns are alike. Likewise the

SAMPLE RELIABILITY OF PER CENT OF FREQUENCY OF VOWEL SOUNDS OF NEWBORNS UNDER THREE CONDITIONS

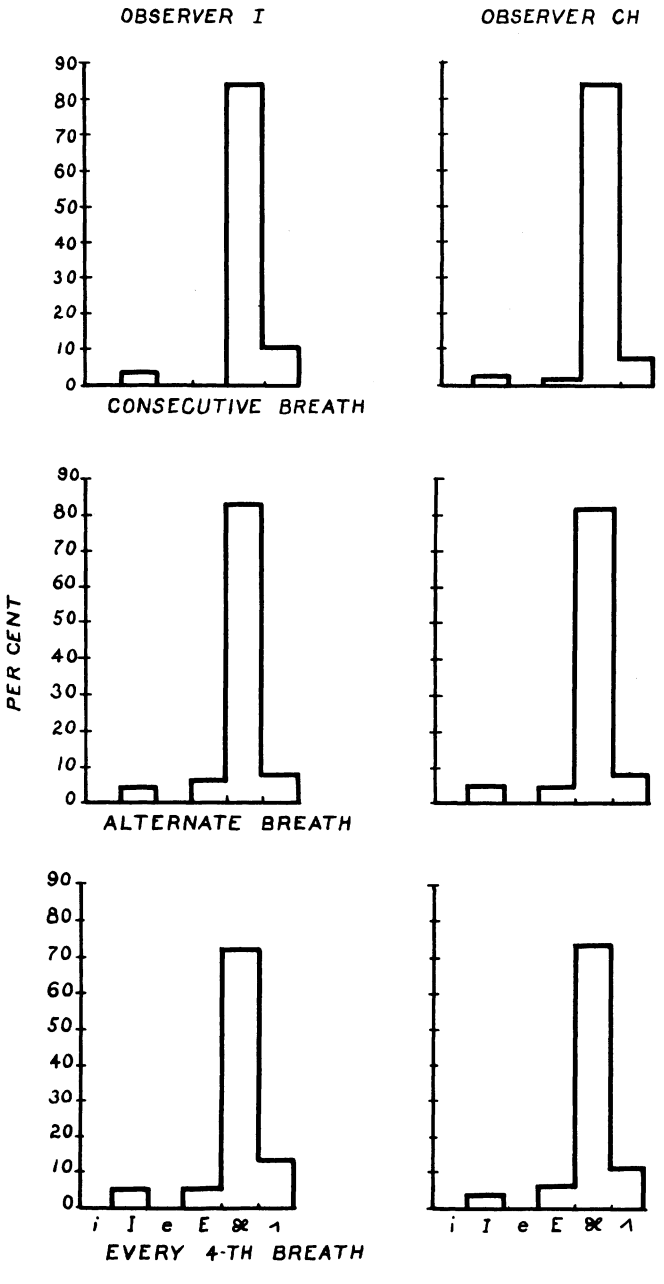

FIGURE 3

three profiles at the right of the figure exhibit a high degree of similarity.

It is readily noted that of the four sounds transcribed in this experiment the vowel æ clearly dominates. Consistently the vowel ʌ maintains second place, while ɪ and ɛ show variation. These vertical comparisons thus demonstrate in a graphical manner the reliability of the data within this sample.

Incidentally a horizontal comparison of the pairs of profiles reveals a high observer reliability within experiment 2.

Comparison of Odd-Even Items

A second demonstration of the reliability of the data is afforded by a comparison of odd-even items. An analysis of these items will be made of the data of both experiments 1 and 2.

The next tabulation presents the odd-even items of different vowels made by the 40 infants in experiment 1. The tabulation is read: for observer I the sound æ occurred 580 times on the odd transcriptions and 585 times on the even, the per cent of agreement being 99. Other sounds are read accordingly.

When the corresponding values for the two observers are averaged and then turned into percentages the results may be seen in the profiles of Figure 4. It is evident from the similarity of these profiles that the reliability of the data in this experiment in terms of odd-even items is quite adequate. In both profiles the æ sound not only dominates the dispersions but their values differ by less than 1 per cent. The order

SAMPLE RELIABILITY OF PER CENT OF FREQUENCY OF VOWELS (40 INFANTS)
EXPERIMENT I.

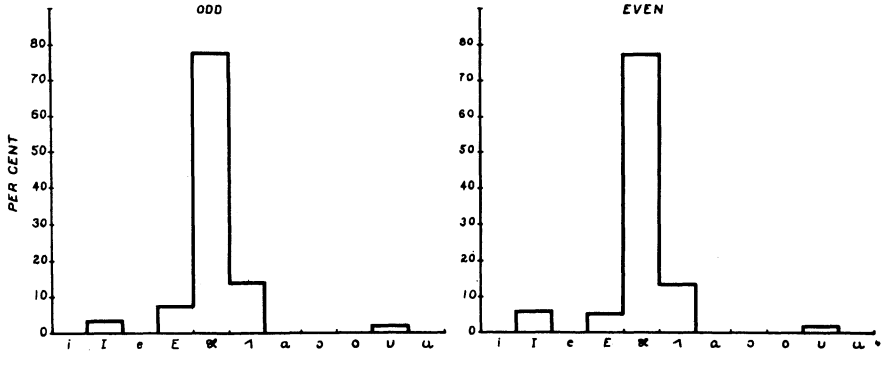

FIGURE 4

of dispersion for the odd and the even values is the same: æ, ʌ, ɛ, ɪ, and ʊ. The corresponding values in each case differ by less than 1 per cent. Separate profiles of the odd-even items of each of the observers have not been included since those of Figure 3, with very slight modifications, could be substituted for them.

In experiment 2, in which five infants were used, the problem of

Vowel	Frequency Odd	Even	Agreement, Per Cent
	Observer I		
I	32	47	67
ɛ	44	41	93
æ	580	585	99
ʌ	109	110	99
ʊ	6	6	100
Total	771	789	
	Observer Ch		
I	19	28	68
ɛ	45	37	82
æ	576	577	100
ʌ	92	95	97
ʊ	6	4	67
Total	738	741	
	Average		
I	25.5	37.5	68
ɛ	44.5	39.0	88
æ	587.0	581.0	99
ʌ	100.5	102.5	98
ʊ	6.0	5.0	83
Total	754.5	765.0	
Average			87

data reliability will be dealt with in two ways: 1) by comparing the odd-even agreement of speech items in each of the categories of consecutive, alternative, and fourth transcriptions; and 2) by comparing these three categories with each other.

Table 2 presents the odd-even frequencies of the different vowels made by five babies under the three different conditions. The table may be read thus: for observer I's data, the vowel sound æ occurred 150 times in his odd consecutive transcriptions, and 151 times in the even ones. The per cent of agreement is 100. This vowel occurred 144 times on his odd alternative transcriptions and 144 on his even ones. It occurred 139 times on his odd fourth transcription and 145 times on his even, etc. The second part of the table gives the corresponding results for observer Ch, and the third part gives the average values for the two observers.

When the averaged results are transformed into per cent values and plotted as profiles the result is seen in Figure 5. The profiles reveal that under the condition of consecutive transcription the order for both the odd and the even values is: æ, ʌ, I, ɛ. For both the alternative and fourth transcriptions the odd and even order is first æ and then ʌ, but there is some variation in regard to the rank of I and ɛ. However the differences between the odd and even values are negligible. The largest difference amounts to 5 per cent between æ recorded on the even transcriptions and æ on the odd transcriptions. This difference is

TABLE 2

FREQUENCIES OF THE DIFFERENT VOWELS MADE BY FIVE BABIES UNDER THE THREE DIFFERENT CONDITIONS

Vowel	Consecutive			Alternative			Fourth		
	Frequency		Agreement, Per Cent	Frequency		Agreement, Per Cent	Frequency		Agreement, Per.cent
	Odd	Even		Odd	Even		Odd	Even	
Observer I									
I	7	9	78	7	5	71	11	11	100
ɛ	0	0		10	10	100	18	8	44
æ	150	151	100	144	144	100	139	145	96
ʌ	24	17	71	16	12	75	29	26	90
Total	181	177		177	171		197	190	
Observer Ch									
I	1	9	11	5	10	50	8	8	100
ɛ	3	5	60	10	9	90	16	9	56
æ	147	147	100	142	145	98	141	143	99
ʌ	12	19	63	20	11	55	24	21	88
Total	163	180		177	175		189	181	
Average									
I	4	9	44	6	7.5	80	9.5	9.5	100
ɛ	1.5	2.5	60	10	9.5	95	17.0	8.5	50
æ	148.5	149	100	143	144.5	99	140.0	144.0	97
ʌ	18	18	100	18	11.5	64	26.5	23.5	89
Total	172	178.5		177	173.0		193.0	185.5	

not significant. Thus here, too, the profiles are similar, indicating that the reliability in terms of the odd-even frequencies within each of the three categories is quite adequate.

Another way of indicating reliability of the data collected on the 5 infants is by a direct comparison of the results of the three conditions themselves. These results are found in the two tabulations and in the profiles.

The first of the two tabulations presents the frequencies of each of the vowels under the conditions of consecutive, alternative and fourth

Vowel	Frequency		
	Consecutive	Alternative	Fourth
Observer I			
I	16	12	22
ɛ	0	20	26
æ	301	288	284
ʌ	41	28	55
Total	358	348	387
Observer Ch			
I	10	15	16
ɛ	8	19	25
æ	294	287	284
ʌ	31	31	45
Total	343	352	370

SAMPLE RELIABILITY OF PER CENT OF FREQUENCY OF SOUNDS OF FIVE NEWBORNS FOR ODD AND EVEN BREATHS UNDER THREE CONDITIONS (BASED ON THE AVERAGED RESULTS OF 1 (CH.)

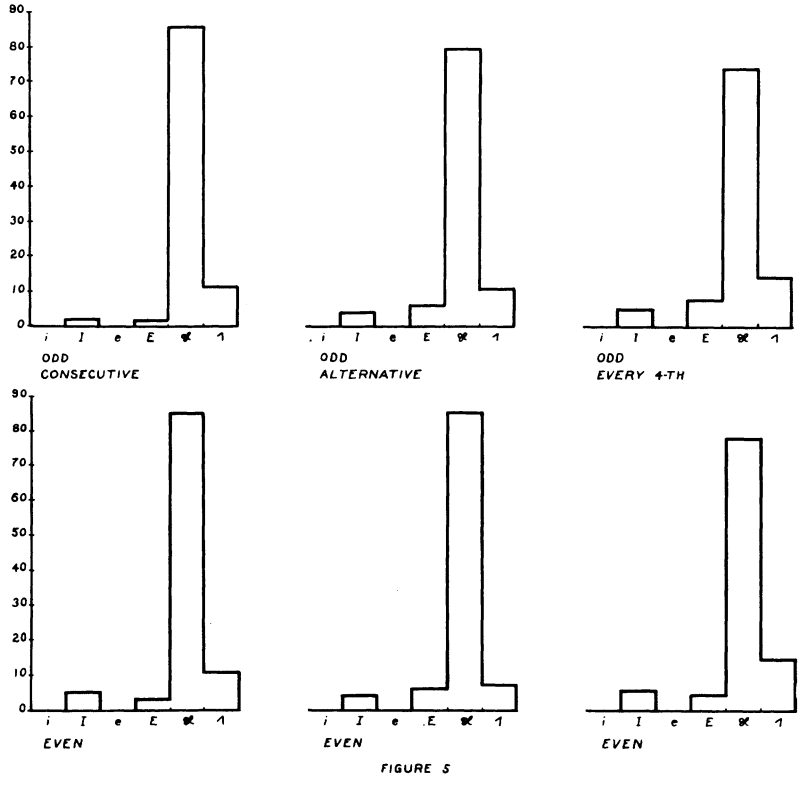

FIGURE 5

transcriptions for the two observers and also for the average of their results. When these values are turned into percentages and these percentages are compared in terms of consecutive vs. alternate, consecutive vs. fourth, and alternative vs. fourth, the results may be seen in

Vowel	Frequency		
	Consecutive	Alternative	Fourth
	Average		
I	13	13.5	19
ɛ	4	19.5	25.5
æ	297.5	287.5	284
ʌ	36	29.5	50
Total	350.5	350	378.5

the next tabulation. The indices are quite satisfactory on the sound æ, fairly satisfactory for the sounds ʌ, and I, and in two comparisons not satisfactory for the sound ɛ.

Vowel	Agreement, Per Cent: Consecutive vs. Alternative	Consecutive vs. Every 4th	Alternative vs. Every 4th
	Observer I		
I	77	77	60
ɛ			85
æ	98	87	89
ʌ	70	81	57
	Observer Ch		
I	67	67	100
ɛ	43	34	79
æ	95	89	94
ʌ	97	75	72
	Average		
I	95	74	78
ɛ	20	16	84
æ	97	88	91
ʌ	82	78	63

DISTRIBUTION OF VOWEL SOUNDS OF FIVE INFANTS UNDER THREE CONDITIONS

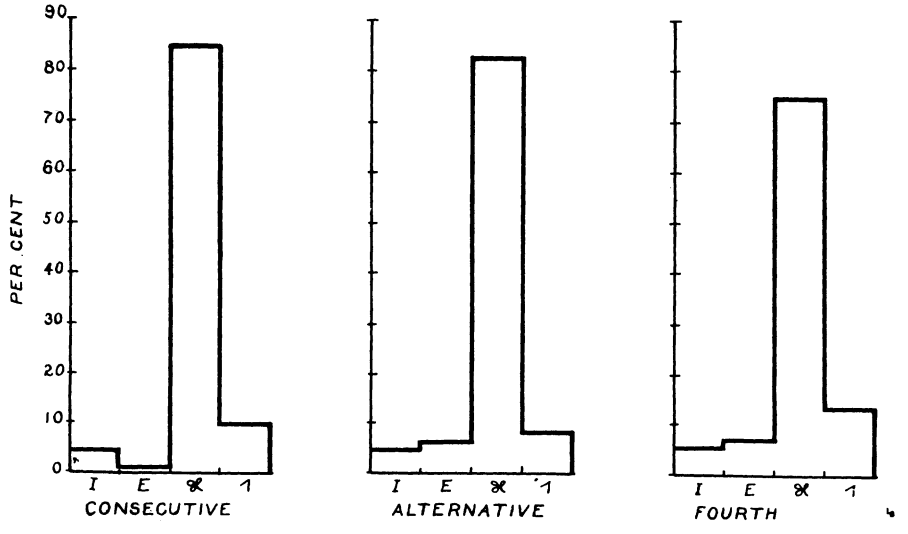

FIGURE 6

When the values in the third part of the frequency tabulation above are turned into per cents and graphed as profiles, the similarity of the patterns becomes evident. The rank order of the sound frequency for æ and for ʌ is first and second under all three conditions. For the alternative and the fourth transcriptions ɛ ranks third, and I fourth. However, in the case of the consecutive transcription ɛ and I are reversed.

Comparison of the Results of Three Experiments

In the discussion of observer reliability the similarity between the profile patterns of pairs of observers in Figure 1 was pointed out. Figure 1 also may be used to graphically illustrate data reliability. By reading the figure vertically it will be seen that the successive profiles of the three samples transcribed by observer I are very similar. The same is true of the two samples collected by observer Ch.

In the distribution of the use of the seven vowels shown in the Irwin-Curry data, the sound æ is strikingly dominant. The ɛ sound ranks second and a few back vowels are present. Likewise in the dispersion of sounds illustrated in the profiles of experiment 1, the element æ is clearly dominant among the five sounds transcribed, and ʌ ranks second. One back vowel is rarely present. Forty infants were used in both of these samples. In the data of experiment 2 on 5 infants, the sound æ likewise stands out prominently, the ʌ sound ranks second, but no back vowels were heard.

Thus Figure 1 may be taken as a demonstration not only of observer reliability, but also of data reliability.

SUMMARY

In order to determine the degree of observer and data reliability which can be achieved on speech sound data, two experiments were conducted on the speech sounds of infants under 10 days of age. The first experiment included a sample of vowels occurring on 30 respirations of each of 40 infants; the second, a sample of 180 respirations on each of 5 infants. Transcriptions were made in the International Phonetic Alphabet by two observers working simultaneously and independently. These data together with those of a previous study constitute the materials which have been analyzed. Methods of calculating the reliability of this type of data are described. The following results were obtained:

1. Whining and crying may be reliably discriminated. The agreement index for the two observers averages 97 per cent.

2. Agreement of observers on crying sounds of the group of 40 infants is 93 per cent on a frequency criterion; for a breath criterion, it is 91 per cent. The range in both cases is from 74 to 100 per cent.

3. On the frequency criterion there was observer agreement above 85 per cent in 36 of the 40 infants.

4. On the breath criterion there was agreement above 85 per cent on 33 of the 40 cases.

5. Profiles of vowel dispersion as transcribed by three pairs of observers are strikingly similar.

6. It is evident that training in transcribing infant sounds on about 18 subjects will yield observer reliability above 90 per cent.

7. There are reasons for believing that when the methods of fre-

quency and breath reliability are both used, a more complete result is obtained than when either is used singly.

8. When transcriptions of samples on consecutive, alternative, and fourth sounds are compared as between observers, the frequency reliability is 96 per cent and breath reliability is 95 per cent.

9. For the separate sounds vocalized by the group of 40 babies, the per cent of observer agreement varied from 60 on the sound I to 99 on the sound ɰ. The average for the five vowels uttered by this group is 85 per cent.

10. For the separate vowel sounds of the group of 5 subjects, the per cent of observer agreement varied from 82 on the sound I to 99 on the sound ɰ. The average for four sounds uttered by this group is 89 per cent.

11. When the reliability of the data is analyzed in terms of consecutive, alternative and fourth samplings and presented graphically in profiles of vowel dispersion, a high degree of similarity is evident.

12. A comparison of odd-even items of the vowel sounds of the group of 40 cases shows agreement ranging from 68 to 99 per cent with an average of 87 per cent.

13. A profile comparison of the odd-even items within each of the conditions of consecutive, alternative, and fourth transcriptions shows close similarity.

14. A comparison of the distributions of sounds between consecutive and alternative transcriptions, between consecutive and fourth, and between alternative and fourth transcriptions yields satisfactory reliability on 9 of 12 indices.

15. When the findings of the two experiments of the present study are compared by means of profiles with the results of a previous experiment it is seen that the data show a high degree of similarity.

16. A sample of the sounds of 30 breaths of 40 infants seems to yield more individual sounds than the sample of 180 breaths of 5 subjects.

It is concluded from these findings that quite adequate observer and data reliability can be attained on the speech sounds of newborn infants.

REFERENCES

(1) Irwin, Orvis C.: Research on speech sounds for the first six months of life. Psychol. Bull., 1941, 38, 277-285.

(2) Irwin, Orvis C., and Curry, Thayer: Vowel elements in the crying vocalization of infants under ten days of age. Child Develop., 1941, 12, 99-109.

(3) Irwin, Orvis C., and Krehbiel, Thelma E.: Speech sounds of infants: The fourth, fifth, and sixth months. Unpublished Study.

(4) Irwin, Orvis C.: The profile as a visual device for indicating central tendencies in speech data. Child Develop., 1941, 12, 111-120.

(5) McCarthy, Dorothea: The vocalization of infants. Psychol. Bull., 1929, 26, 625-651.