7900796

MCGARR, NANCY STEINMULLER
THE DIFFERENCES BETWEEN EXPERIENCED AND INEXPERIENCED LISTENERS IN UNDERSTANDING THE SPEECH OF THE DEAF.

CITY UNIVERSITY OF NEW YORK, PH.D., 1978

University Microfilms International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

THE DIFFERENCES BETWEEN EXPERIENCED AND INEXPERIENCED LISTENERS IN UNDERSTANDING THE SPEECH OF THE DEAF

by

NANCY STEINMULLER McGARR

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1978

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

August 21, 1978
date

Harry Levitt
Chairman of Examining Committee

August 21, 1978
date

Irving Hochberg
Executive Officer

Katherine S. Harris

Irving Hochberg

Harry Levitt
Supervisory Committee

Joanne Subtelny
Consultant

The City University of New York

TABLE OF CONTENTS

Page

Page

ACKNOWLEDGEMENTS

All good things must come to an end--and fortunately also dissertations! This section appropriately headed "Acknowledgements" indicates that I have finally arrived at the day when it is not presumptuous to extend my thanks to those who have assisted in this endeavor. It is altogether appropriate that these words, among the last written, appear first in this dissertation. They are, in the writer's opinion, the most important words in this manuscript.

One of the most extraordinary aspects of this program in Speech and Hearing Sciences is the people who are part of it. The faculty, the staff, and my fellow students have contributed to make these years a richly rewarding personal and professional experience.

Professors Harry Levitt and Katherine Harris have guided my academic studies, and this thesis, with wisdom, good humor, and unending support. It is my privilege to be counted as one of their students. They are masters at their art.

I am also grateful to Professor Irving Hochberg who served most graciously on this dissertation committee.

My colleagues in Room 904 provided immeasureable good will and encouragement as this work progressed. It was their spirit of friendship that often kept the work progressing at all.

Without the cooperation of over one hundred fifty listeners, the differences between experienced and inexperienced listeners would still be mere conjecture. For their words of encouragement and skilled listening, I am indebted to the faculty and staff of: the Cleary School for the Deaf--Western Campus, Gallaudet College, the Lexington School for the Deaf, the Mill Neck Manor School for the Deaf, and the National Technical Institute for the Deaf.

Thanks are also due to my inexperienced listeners: my family and friends, and especially my students at Molloy Catholic College for Women who volunteered most willingly.

Mr. Harvey Stromberg's contribution to this data analysis was highly significant. His computer skills and superb sense of humor were equally indispensable to completing this work.

I am also grateful to Haskins Laboratories for graciously providing me with computer time to prepare the speech stimuli for this project.

My family lent their ears not only to this study, but also to my extended discourse on the progression of this thesis. They were understanding, patient, and encouraging when I most needed them. Any credit for arriving at this day must truly be shared with them.

But if there is one person who knows as well as I, the hours, the weeks, and the years that have been spent in pursuing this degree, it is John McGarr. For his belief in me, and in my dream, I have no adequate words at all.

To each of you, my prayerful thanks.

N.S.M.
Rye, New York
August 1, 1978

LIST OF TABLES

Page

LIST OF FIGURES

Figure Page

Figure		Page

CHAPTER I

INTRODUCTION

To most educators of the deaf, it would come as no surprise that the deaf child whose speech seems intelligible in the classroom may be virtually unintelligible to the "man on the street." Several investigations (Brannon, 1964; Markides, 1970; Smith, 1972) have noted the difficulty that naive listeners have in understanding the speech of the deaf. The percent deaf speech intelligible to naive listeners has been reported as ranging from 18.7% (Smith, 1972) to 20.7% (Brannon, 1964).

In contrast, the experienced listener's ability to understand deaf speech is clearly superior to that of the naive listener (Mangan, 1961; Thomas, 1963; Markides, 1970; Monsen, 1978). The fact that an experienced listener understands deaf speech with greater accuracy than inexperienced listeners has been assumed not to be related to his personal knowledge of the speaker. It would seem that once familiar with deaf speech, the experienced listener is able to generate higher intelligibility scores for deaf speakers in general. Indeed, the difference between experienced and inexperienced listeners is so widely accepted that the intelligibility of deaf speech has been rated according to how likely the speaker may be understood by "most trained teachers of the deaf, most people familiar with deaf speech, or almost everyone." (Thomas, 1963).

Calvert (1961) noted that experienced listeners could consistently identify a speaker as deaf only when there was some articulatory movement in the speech. As the amount of articulatory movement increased (e.g. from a vowel, to a CVC nonsense syllable, to a sentence), the listener's ability to identify the deaf speaker also increased. When the speakers produced sentences, the listeners were able to identify which samples were produced by the deaf speakers one hundred percent of the time. These results suggest that certain aspects of speech production in the deaf provide the listener with acoustic information that is quite specific to the deaf population and which can be readily distinguished from normal production.

Several investigators (Hudgins and Numbers, 1942; Markides, 1970; Smith, 1972) have been remarkably consistent in identifying typical articulatory errors in the speech of the deaf. Briefly, these include: omissions, substitutions and severe distortions of the intended phoneme. Among the more common error types are: incorrect production of palatals, alveolar fricatives, affricates and the nasal velar. In general, this work suggests that as the frequency of segmental errors increases, overall intelligibility decreases.

Other investigations have also identified errors in suprasegmental production which are characteristic of the speech of the deaf, although the extent to which these effect overall intelligibility is less well established. These included: errors in timing such as excessively slow rate (Calvert, 1961; Brannon, 1964), increased duration of phonemes (Calvert, 1961; John and Howarth, 1964; Monsen, 1974; Osberger, 1978), and inappropriate pause (Levitt et al., 1974; Osberger, 1978; Parkhurst and Levitt, 1978). In addition, deviation in phonatory control including inappropriate pitch (Angelocci et al., 1964; Levitt et al., 1974), excessive variability in pitch (Levitt et al., 1974; McGarr and Osberger, 1978), and less than normal variation in pitch resulting in flat and monotonous speech (Green, 1956; Hood, 1966; Smith, 1972), also contribute to reduced intelligibility.

The consistency of the reported error patterns in segmental and suprasegmental production suggests that one possible explanation for the differences between experienced and inexperienced listeners may be found in the ability of the experienced listeners to recode deaf speech so as to compensate for typical errors in production.

Another factor of importance may be the relative ability of the experienced listeners to make use of contextual information. Hudgins and Numbers (1942) suggested that the inexperienced listener was so distracted by the quality of the speech of the deaf, that she could not derive information from contextual cues as could the experienced listener. On the other hand, the difference in scores between isolated words and sentences led Brannon (1964) to conclude that context was extremely important for naive listeners. Similarly, Thomas (1963) showed that scores for "everyday" sentences were higher than those for isolated monosyllabic words for both experienced and naive listeners.

In these investigations, and in others as well (Hudgins, 1949; Subtelny, 1977), context was defined as words produced and heard in sentences. While the studies concluded that words produced and heard in sentences received higher scores than words produced in isolation (for either experienced or inexperienced listeners), specification of context *per se* was not made. Several factors may be playing a part such as the context of the sentence as well as differences in articulation.

Thus, it is not known if the differences in scores for isolated words and sentences may be attributed to (1) contextual constraints in the sentences, (2) the articulatory environment; that is, the articulation of words in isolation is not identical to those in a sentence, or (3) an interaction of these. Furthermore, it is not known to what extent the experienced and inexperienced listeners may use context to disambiguate errors in production.

Although there appears to be a significant difference between experienced and inexperienced listeners' ability to understand the speech of the deaf, a systematic investigation to isolate those factors which contribute to the differences between groups has not been conducted.

The purpose of this study was to isolate several factors which were believed to contribute to the reported differences between listeners.

Specifically, the study was designed to answer the following questions:

1. Is there a significant difference between experienced and inexperienced listeners in understanding the speech of the deaf?

2. What is the effect of listener experience on the intelligibility of words produced in isolation, in sentence context, and produced in sentence context but heard in isolation?

3. What is the effect between listener experience and context, e.g. does context benefit the experienced listener more than the inexperienced listener?

4. Are any of the above effects influenced by the relative intelligibility of the test words?

CHAPTER II

REVIEW OF THE LITERATURE

The Differences Between Experienced and Inexperienced Listeners

Hudgins and Numbers (1942) in their classic study on the speech of the deaf claimed there was a significant difference between experienced and inexperienced listeners. Experienced listeners were thought to use sentence context in addition to articulatory information obtained from viewing the speaker's face to increase intelligibility scores. In fact, intelligibility scores obtained by experienced listeners were 10-15% higher when they were permitted to see the speaker's face than when only auditory cues were permitted. On the other hand, Hudgins and Numbers suggested that the inexperienced listener was so distracted with the unusual voice quality of deaf speakers, that overall speech intelligibility was severely impaired. However, data was not collected on intelligibility scores for inexperienced listeners. Nevertheless, Hudgins and Numbers reasoned that the inexperienced listener was of little value as an evaluator of deaf speech.

Data on experienced and inexperienced listeners were obtained in a study by Thomas (1963). Three groups of listeners (6 listeners in each group) were tested on their ability to understand deaf speakers. The three groups were composed of: (1) naive listeners--no previous exposure to the speech of the deaf, (2) semi-naive listeners--college seniors enrolled in a teacher training program who had worked with the deaf for approximately 1 year, and (3) experienced listeners--teachers of the deaf.

A group of 8 deaf speakers (pure tone averages unspecified) produced a list of 50 words from the CID Auditory Test W-22 as well as a list of 10 CID sentences of "everyday American speech" also known as the CHABA sentences (Davis and Silverman, 1978). Half of the listeners in each group evaluated the 8 deaf speakers using only auditory cues to transcribe speech production; the remaining half of the listeners evaluated the 8 deaf speakers using combined auditory and visual cues.

The overall results showed that scores increased from naive to experienced listeners with the largest intergroup differences occurring between the naive and semi-naive groups. Table 1 shows the range and mean scores for sentences and words in the two listening conditions and across the three levels of listener experience. Several observations can be made in this data:

1. The experienced listeners' scores were consistently higher than those of the semi-naive or naive listeners for sentences as well as for isolated words in either listening condition. There was also considerable overlap in scores between groups of listeners for sentences or words. However,

TABLE 1

PERCENT INTELLIGIBILITY OF WORDS AND SENTENCES OBTAINED BY EXPERIENCED, SEMI-NAIVE, AND NAIVE LISTENERS (THOMAS, 1963)

Listeners	Auditory Cues				Auditory and Visual Cues			
	Words		Sentences		Words		Sentences	
	Mean Percent	Range of Percent	Mean Percent	Range of Percent	Mean Percent	Range of Percent	Mean Percent	Range of Percent
Naive	27.5	12.7-48	50.3	18.1-80	38.3	29.3-58.9	68.4	36.1-88
Semi-Naive	34.9	24-54	69.5	41.6-84.7	52.0	38.0-68.7	88.5	76.7-97.4
Experienced	37.0	22-52	76.8	51.1-95.5	59.0	44-74	92.0	80.8-100

as a group, the scores of the naive listeners did not approach those of the experienced listeners.

2. Scores for sentences were consistently higher than those for words by all listeners. In fact, the percent intelligibility for the "everyday sentences" was almost double that of the monosyllables.

3. Both experienced and inexperienced listeners obtained scores 10-22% higher for the combined auditory-visual condition than for the auditory-only condition. This improvement in intelligibility scores when visual cues were permitted was similar to the gain reported by Hudgins and Numbers (1942).

The results of Thomas's study showed that experienced listeners were clearly superior to the semi-naive or naive listeners in understanding the speech of the deaf. However, all listeners obtained higher scores for sentences than for words, and also for the auditory-visual condition than for the auditory-only condition. These results suggest that effective use of sentence context and visual information is not limited to the experienced listener as proposed by Hudgins and Numbers.

The Use of Experienced and Inexperienced Listeners to Evaluate the Intelligibility of Deaf Speech

Numerous investigators (Mangan, 1961; Brannon, 1964; John and Howarth, 1965; Markides, 1970; Smith, 1972; among others) have used experienced or inexperienced listeners to obtain measurements of the intelligibility of deaf speech.

Mangan (1961) used teachers of the deaf, mothers of deaf children, and naive college students to evaluate a group of partially deaf children (mean hearing loss 65-75 dB) and a group of deaf children (mean hearing loss greater than 75 dB). Each child read a list of 50 PB words before the panel of listeners who were allowed to watch and listen to the children when transcribing the speech production samples. The percent of words correct was computed for the three groups of listeners and averaged across all hearing impaired speakers. The mean score for each group of listeners was: experienced listeners--65%, mothers of deaf children--51%, and inexperienced college students--50%. No significant difference was found between the parent group and the inexperienced college students but a significant difference was found between the parent group and the teachers. These intelligibility scores for PB words are slightly higher than the 59% for experienced and 38% for naive listeners reported by Thomas (1963), but are approximately the same for the semi-naive listeners, in both studies.

Brannon (1964) used naive listeners to evaluate the speech production of severe to profoundly deaf children. The speech stimuli included 50 PB words, 10 spondees, 10 polysyllabic words, ten 2-word sentences, ten 4-word sentences and 10 sentences from the CID sentences of "everyday American speech." The naive listeners transcribed the speech stimuli on a word-by-word basis using only auditory cues. The scores for overall intelligibility ranged from 7.8% to 35.6% with a mean score of 20.7%. Scores for sentences were

26% while those for words in isolation were approximately 17%. While this trend is similar to that reported by Thomas (1963), overall intelligibility scores obtained by Brannon's naive listeners were considerably poorer for PB words. In general, as the number of words in the sentence increased, the overall intelligibility increased. However, an increase in intelligibility was not observed as the number of syllables in the isolated words increased. Since scores for sentences were higher than those for isolated words, and also intelligibility scores increased slightly as the overall sentence length increased, Brannon concluded that sentence context was important for the naive listener.

John and Howarth (1965) assessed the importance of timing errors on the overall intelligibility of deaf speech. Experienced teachers of the deaf and university students with no previous experience with the hearing impaired, judged the intelligibility of sentences produced by deaf children before and after a training program. The first set of sentences was spontaneously generated by the children in response to picture stimuli. This set constituted the before-training sample. Each child then practiced the sentences he had produced, imitating a teacher-model's rate of speaking. These samples subsequently constituted the after-training stimuli. The authors claimed that there was a significant difference between the intelligibility scores of the experienced and inexperienced listeners but reported only the inexperienced listeners' scores. These were: 19% for sentences produced before training, and 30% for sentences produced after training. The overall intelligibility scores for sentences in the before-training condition is very similar to the 20% score obtained by Brannon's naive listeners.

Markides (1970) used experienced teachers of the deaf and university students with no previous experience, to transcribe the speech production of hearing impaired speakers on a word-by-word basis and to rate the speech for overall intelligibility. The subjects were a group of partially deaf children (mean hearing loss 57 dB) and a group of deaf children (mean hearing loss 95 dB) who described several picture stimili. The experienced listeners obtained higher scores and rated the speech more intelligible than the inexperienced listeners. However, considerable overlap in scores was apparent especially for the better hearing impaired speakers. The mean percent intelligibility scores for the partially deaf children were: 83% for the experienced and 76% for the inexperienced listeners; for the deaf children, 31% for the experienced and 19% for the inexperienced listeners. A correlation of .94 was obtained for the intelligibility scores between the experienced and inexperienced listeners. Markides noted that even when the scores of the inexperienced listeners were 50% or better, the speech was reported as difficult to follow.

Smith (1972) obtained measurements of intelligibility of 20 congenitally deaf children using large groups of inexperienced listeners. Each child read a list of sentences designed to evaluate the production of vowels and consonants in key words. The inexperienced listeners were asked to transcribe the sentences on a word-by-word basis using only auditory cues and to rate the production for overall intelligibility. The intelligibility scores for the naive listeners ranged from 0-76% with a mean of 18.7%; average ratings were not reported. However, Smith noted that overall intelligibility ratings were not always consistent with the number of words understood correctly. Inexperienced listeners frequently rated the child's speech intelligible, when in fact, they had understood very few words correctly. However, the correlation between the ratings and the intelligibility scores was high (p= -.83). It is interesting to note that the intelligibility scores obtained by naive listeners in this study, and in those by Brannon (1964), John and Howarth (1965) and Markides (1970), are all remarkably close to 20%, but considerably poorer than those reported by Thomas (1963) for naive listeners.

Nickerson (1973) used a group of listeners highly experienced with deaf speakers and a group of listeners with little or no experience to assess intelligibility. Six deaf children recorded read and spontaneous speech samples as part of a study to evaluate speech before and after training. For each child, scores could range from 0-100% representing a continuum of poor to good performance. The data for experienced and inexperienced listeners are summarized in Table 2.

TABLE 2

PERCENT INTELLIGIBILITY OF READ AND SPONTANEOUS SPEECH FOR EXPERIENCED AND NAIVE LISTENERS (NICKERSON, 1973)

	Read Speech		Spontaneous Speech	
	Before	After	Before	After
Experienced				
Mean	59%	73.5%	47.5%	59.8%
Range	22-98%	18-99%	17-95%	26-98%
Naive				
Mean	34.1%	36%	27.6%	24.3%
Range	4-91%	3-88%	0-84%	3-58%

Briefly, Nickerson's data showed: (1) scores for experienced listeners were consistently higher than those for naive listeners both for read and spontaneous speech, (2) all listeners obtained higher scores for read speech than for spontaneous speech, (3) the difference between read and spontaneous speech scores was greater for the experienced listeners, and (4) experienced listeners scored samples recorded after training higher than those in the before-training condition for most all children. However, there was little difference between the intelligibility scores for samples before or after training as evaluated by the naive listeners. In many instances, the naive listeners scored the speech samples produced after training, lower than the samples produced before training.

A scatterplot was obtained of the intelligibility scores for read and spontaneous samples obtained by experienced and naive listeners. Although intelligibility scores for experienced listeners were higher than those of naive listeners, a linear relationship was found between the groups. Nickerson suggested that the difference between experienced and naive listeners was relatively stable across deaf speakers with a wide range of intelligibility. The data suggests that when the experienced listeners' scores fell within the range of 40-80%, the naive listeners' scores would be expected to be 30-40% lower.

Monsen (1978) used 35 naive and 15 experienced listeners to assess the intelligibility of a large group of deaf students who produced simple sentences and questions. The speech materials contained common monosyllables and spondees so that even the poorest speaker was partially intelligible. Each listener heard four successive tapes containing 39 sentences spoken by the deaf subjects. The listeners transcribed the sentences on a word-by-word basis. Percent intelligibility was weighted for each sentence with common function words assigned a smaller value than less common content words.

Naive listeners received lower scores than experienced listeners. The mean score for the naive listeners across the four listening tapes was approximately 72.5%; for experienced listeners, the mean score was approximately 83%. For the first tape heard, the difference between the mean of the two listening groups was approximately 14%, but the difference between listeners decreased as the amount of listening time increased. By tape four, the difference between the mean scores for the groups was only 5%. Considerable overlap in scores for the last three tapes was also noted. However, this may be an effect of the scores approaching 100%. Monsen suggested that the advantage of experience in listening to the speech of the deaf was one easily acquired and in a relatively short period of time.

Table 3 summarizes intelligibility scores obtained by experienced and inexperienced listeners as reported in the literature. Several observations may be made in these data.

1. A difference in intelligibility scores obtained

TABLE 3

INTELLIGIBILITY SCORES OBTAINED BY EXPERIENCED AND INEXPERIENCED LISTENERS IN STUDIES ON THE SPEECH OF THE DEAF

Study	Percent Intelligibility For Experienced	Percent Intelligibility For Inexperienced
Isolated Words		
Hudgins, 1949.	42%	----
Mangan, 1961	65% *	50% *
Thomas, 1963	37%	28%
	59%	38%
Brannon, 1964.	----	17%
Subtelny, 1977	35%	----
Sentences		
Hudgins, 1949.	71%	----
Thomas, 1963	59%	50%
	92%	68%
Brannon, 1964.	----	26%
John and Howarth, 1965	----	19%
Markides, 1970	31%	19%
Smith, 1972.	----	18.7%
Nickerson, 1973.	47% speech	28% speech
	59% reading	34% reading
Subtelny, 1977	66%	----
Monsen, 1978	83%	73%

*Auditory and Visual Cues

by experienced and inexperienced listeners has been consistently reported in several investigations (Mangan, 1961; Thomas, 1963; Nickerson, 1973; Monsen, 1978). Overall intelligibility scores decreased from experienced to semi-naive to naive listeners. While there may be considerable overlap in scores for individual listeners, as a group the scores of the naive listeners did not approach those of the experienced listeners.

2. Scores obtained by experienced and inexperienced listeners were higher for sentences than for isolated words (Hudgins, 1949; Thomas, 1963; Subtelny, 1977).

3. Both experienced and inexperienced listeners evidenced a wide range of intelligibility scores for sentence material. For example, scores for inexperienced listeners ranged from 18.7% (Smith, 1972) to 73% (Monsen, 1978); scores obtained by experienced listeners ranged from 31% (Markides, 1970) to 83% (Monsen, 1978).

This wide variation in intelligibility scores for either group of listeners may be related to the speech stimuli the deaf subjects were asked to produce. Scores which clustered near the low end of the intelligibility range for either experienced or inexperienced listeners were generally obtained from sentence material such as spontaneously generated speech samples (John and Howarth, 1965; Markides, 1970; Nickerson,1973). Scores which clustered near the high end of the intelligibility range for either experienced or inexperienced listeners were those which contained words of common usage or were highly redundant such as the CID sentences of "everyday American speech" (Thomas, 1963; Monsen, 1978). However, the magnitude of difficulty appears to be similar for both groups; that is, speech materials devoid of context seemed to be considerably more difficult for both groups of listeners than those high in context.

The Effect of Training Listeners on Intelligibility Scores

The difference between listeners in understanding the speech of the deaf raises important questions in evaluation procedures. What are the effects of training listeners prior to evaluation? How much listening time must elapse during the evaluation before the scores of the inexperienced listeners approach those of the experienced? These questions have received relatively little attention in the literature on the deaf.

According to Hudgins (1949) use of either experienced or inexperienced listeners to judge the intelligibility of deaf speakers may result in overall scores that are too high or too low, therefore not accurately representing the intelligibility of the deaf speaker. Hudgins suggested that evaluations should only be conducted by a trained panel of listeners.

Inexperienced undergraduates were selected to form a panel and were trained on listening tapes prior to evaluations. The training tapes consisted of 8 deaf speakers with a wide range of intelligibility, each of whom produced a list of 50 phonetically balanced monosyllables. A training session lasted 1 1/2 to 2 hours during which time the listeners heard the entire set of recordings produced by the 8 deaf speakers played twice. Listeners' scores were computed as a percentage of words correctly understood at each session and learning effects were noted. By the end of the 6th training session (approximately 12 hours), the listeners had reached a plateau in their learning. The most dramatic gain in scores occurred during the first three sessions (approximately 6 hours). However, it should be noted that the same children reading randomized versions of the same word list were heard during each training session.

Improvement in scores was also noted after the training sessions were completed. The listening panel evaluated the speech of a large group of deaf children who produced the lists of 50 monosyllables and a list of sentences. The 8 deaf subjects from the training sessions were also included in this evaluation. During the test procedure, the 8 subjects trained on received intelligibility scores 10% higher than had been awarded them during the last training session. Since the speech samples for these 8 speakers were the same for both training and evaluation, the gain in intelligibility could not be accounted for by improvement on the part of the speakers. Rather, Hudgins held that improvement in scores was due to familiarity with the test material and the increased experience obtained by the listeners in hearing many deaf speakers. Hudgins showed that an inexperienced listener's ability to produce higher intelligibility scores may be the result of training prior to evaluation (which may include familiarization with the speech as well as the material), and also of learning effects which occurred during the testing.

Apart from Hudgins, only Monsen (1978) has shown that the scores of inexperienced listeners will improve while listening to the speech of deaf children. Briefly, as the amount of listening time increased (from tape 1 to tape 4), the difference between experienced and inexperienced listeners decreased. Also intelligibility scores showed considerable overlap. The difference between listeners' average scores on the first tape was 14%, but only 5% on the last tape.

While only Hudgins (1949) and Monsen (1978) have data on learning effects which occurred during evaluations, and only Hudgins , on the effects of training listeners prior to evaluation, the effect of listener experience has been acknowledged in the literature in a somewhat general way.

For example, Mangan (1961) used experienced teachers, mothers of deaf children and inexperienced listeners to assess the intelligibility of deaf speakers. Mangan assumed that mothers of deaf children had more experience than naive listeners but somewhat less experience than classroom teachers. Thus, the percent intelligibility of deaf speech of these three groups would be a function of their experience with the deaf. Briefly, scores obtained by the mothers were essentially the same as the inexperienced listeners. These scores in turn were considerably poorer than those of the experienced teachers.

On the other hand, Thomas (1963) showed a difference in intelligibility as a function of listener experience. Semi-naive listeners, defined as students in teacher training program with approximately one year's work with the deaf, achieved scores that were poorer than the experienced teachers of the deaf, but considerably better than the inexperienced listeners. Thomas suggested that the greatest gain in intelligibility occurs during the first year of exposure to deaf speech with smaller gains thereafter.

That improvement in the scores of listeners with little or no experience with the deaf could be affected by training prior to evaluation or learning effects which occur during evaluation is not surprising. In the speech perception literature, attention has been given to listener effects in evaluating the intelligibility of speech. Briefly, this work has shown that providing the listener with training prior to listening, or more specifically training the listener on the test materials, will improve performance.

Egan (1948) found that listeners who received training prior to evaluating speech-in-noise received higher scores than untrained listeners. Significant improvements in the scores of the untrained listeners were shown with practice. Stable scores were obtained after 10-15 hours of practice over a 5-7 day period. However, the nature of the training given the experienced listeners prior to the evaluation was not specified. The experienced listeners may have been trained on the speech materials *per se* or participated in other speech-in-noise evaluations, or both. Similarly, the improvement in the inexperienced listeners' scores was not quantified.

Improvement in scores was also noted by Moser and Dreher (1955). Stable listening scores for speech-in-noise testing were obtained by inexperienced listeners in 8-10 hours over a 4-5 day period. Also familiarizing the listeners with the message set in quiet resulted in higher scores, in subsequent evaluations. A correlation of .65, significant at the .05 level was obtained between listener performance and training.

Milner (1973) also suggested that listener effects could be controlled if a highly reliable panel of listeners were trained on the test materials prior to evaluations.

In sum, the experience of the listener, training prior to evaluation, and learning effects during testing, may all influence intelligibility scores of deaf speakers. Most studies evaluating the speech of the deaf have used large groups of listeners, frequently inexperienced, to evaluate intelligibility. These large groups served to reduce variability. Furthermore, the amount of speech each listener heard was usually limited in order to avoid problems such as those which might occur with increased familiarity of the test material or the speaker.

Coincidentally, studies by Hudgins (1949), and by

Egan (1948) and Moser and Dreher (1955), have shown that plateaus in the learning curves for deaf speech and for speech-in-noise testing respectively, occurred after 10-12 hours of exposure. In fact, Thomas (1963) has suggested that deaf speech imposes a "hearing loss" on listeners with normal hearing that is analogous to adding noise to the speech signal. This discrimination loss, according to Thomas, is greater for listeners with little or no experience with the deaf. However, systematic studies to substantiate this notion have not been conducted. The combination of segmental and suprasegmental errors in production as well as the substantial language problems of the deaf would suggest that the intelligibility of deaf speech, especially when produced spontaneously, may not be the simple discrimination problem that Thomas suggests.

The Effect of Context

While it is clear that listener experience, training, and learning effects during testing, may influence the intelligibility scores of deaf speakers, contextual factors may also be highly important. In the literature on the deaf, the effect of sentence context has been noted primarily as informal comments. For example, Hudgins and Numbers(1942) suggested that cues a listener obtained from sentence context would limit the effect of articulatory errors on intelligibility. However, a detailed investigation of the effect of sentence context on such errors in deaf speech is non-existant. Studies such as Hudgins (1949), Thomas (1963), Brannon (1964) and Subtelny (1977) have shown that words produced and heard in sentence context received higher intelligibility scores than words produced and heard in isolation. However, these studies have not investigated the effect of context on articulation *per se* (e.g. the same word in isolation and in sentence context).

Hudgins (1949) compared the intelligibility scores for phonetically balanced monosyllables and sentences produced by deaf children. Listeners who had been trained to evaluate the speech production of the deaf transcribed the recordings. Scores for the production of monosyllables ranged from 9-86% with a mean score of 42%. Scores for the production of sentences were 35-100% with a mean of 71%. Over 30% of the children achieved scores of 90% or better on the sentence lists, although few were characterized as having perfectly intelligible speech.

When the scores for sentences were plotted against those for words, a large proportion of the sentences clustered near the high end of the scale. That is, a sentence score of 90-100% corresponded to a word score of approximately 50%. Correlations of .76 (Pearson Moment Method) and .81 (Correlation Ratio Method) were obtained between words and sentences. According to Hudgins, sentence lists may be less discriminating over a wide range of speakers. On the other hand, Hudgins also felt that word lists may underestimate the intelligibility of the speaker, since a word produced in isolation does not convey contextual cues.

As shown in Table 1, Thomas (1963) also compared the intelligibility scores for monosyllables and sentences obtained for naive, semi-naive, and experienced listeners. Scores for sentences were almost double those of monosyllables for each group of listeners. The difference between words and sentences increased from naive to experienced listeners suggesting that experience enabled the listener to take greater advantage of contextual cues and thereby improve overall intelligibility.

Subtelny (1977) also investigated the difference in intelligibility of isolated words (W-22 lists) and sentences (Clark School for the Deaf). The stimuli were produced by young deaf adults and 5 experienced listeners transcribed the recorded speech. Scores obtained for words produced in isolation by 66 deaf subjects were 35.11% (standard deviation of 23.17%). Scores for sentences produced by 156 deaf subjects were 66.13% (standard deviation of 30.00). As in Thomas's data, the advantage of sentence context resulted in scores nearly double those for isolated words. Subtelny suggested that any measurement of overall speech intelligibility should be made on the basis of speech produced in context.

The advantage of context to increase the intelligibility of deaf speech parallels results obtained in speech perception literature (Miller et al., 1951; O'Neill, 1958; Lieberman, 1963; Pollack and Pickett, 1963; among others). In general, these studies have consistently demonstrated that words heard in sentence context received higher intelligibility scores than did isolated words. However, these studies differ from those with deaf speakers, since the isolated words in this literature were originally produced in sentences, segmented from context, and presented to the listener in isolation.

Miller, Heise and Lichten (1951) examined the effect of context on intelligibility. Test words were produced in sentences and presented to the listener in context. The same test words were also segmented from the sentence and presented to the listeners in isolation. The words produced in sentence context and the segmented words were heard by listeners in a wide range of signal-to-noise ratios. Briefly, scores for test words produced in sentences were consistently higher than those for segmented words. As the signal-to-noise ratio increased, the difference in intelligibility between words and sentences increased. Also limiting the corpus of vocabulary improved intelligibility scores. The authors concluded that producing a test word in context was analogous to limiting the listener's vocabulary since both factors limited the possible alternatives available to the listener and thereby resulted in improved intelligibility.

O'Neill (1957) replicated Miller et al. (1951) using a larger corpus of words presented in sentence context and also in isolation. Once again, scores for words in sentence context were higher than for segmented words. As the signal-to-noise ratio decreased, the differences in intelligibility scores between words in sentences and segmented words also decreased. However, as the signal-to-noise ratio became more favorable, sentence context contributed at least a 20% increase in intelligibility scores above those for segmented words. This percent gain in intelligibility is consistent with that found by Miller et al. (1951).

In the previous investigation, context was defined as a word produced in a sentence. But the degree of sentence context was not controlled. Lieberman (1963) examined sentence context, defined as the amount of redundant (high context) or non-redundant (low context) information available to the listener. The speech stimuli consisted of lists of key words produced in sentences that contained high amounts of contextual information (i.e. familiar maxims and stereotyped phrases) as well as in sentences that contained little context. Each key word was segmented from the sentences using electronic gating techniques. Word boundaries were determined by visually monitoring the speech waveform while simultaneously hearing the signal. The segmented words were re-recorded on listening tapes. The tapes consisted of words produced in sentence context and heard in sentence context, as well as words produced in sentence context but heard in isolation (i.e. segmented words).

Lieberman's results clearly demonstrated the advantage of sentence context in increasing intelligibility. Scores for words heard in context were 100% and consistently higher than the same words in isolation. Segmented words produced in high context sentences were 24% and considerably poorer than the same word produced in a low context sentence. The mean score for low context segmented test words was 59%. The speakers were found to produce the low context words with greater stress, longer duration, or more precise articulation than when the word had appeared in a high context sentence. Lieberman surmised that speakers used different articulatory strategies in producing high or low context sentences, in order to resolve any linguistic ambiguities in the message heard by the listeners.

Pollack and Pickett (1963) hypothesized a trading relationship between differences in production (in this study defined as speaking rate), context (defined as length of utterance) and intelligibility. That is, high intelligibility scores might be achieved by providing the listener with more context (longer utterances) or with more precise articulation of the message (as in a slower speaking rate). Speakers read the materials at three different rates: deliberately slow, normal, and deliberately fast. One or more words were segmented from the passage using gating procedures as Lieberman (1963). Samples varied from 1-8 words in length for each speaking rate and were then re-recorded on listening tapes.

Since intelligibility scores were interpreted for individual speakers, at each speaking rate, and for each sample varying in length, only overall trends will be discussed here. Samples equal in duration received essentially similar scores for any speaking rate. Speech produced at the slower rate did provide the listener with less context but more precise articulation. This condition also provided the listener with more processing time. On the other hand, the faster rates of speaking provided the listener with more context, but generally less accurate articulation. When the utterance varied in length, intelligibility increased as the number of words in the sentences increased. Similar results were also obtained by the authors when the speech stimuli examined were excerpts of conversational speech (1963).

Table 4 summarizes the intelligibility scores for isolated words, segmented words, and sentences reported in the literature. The effect of sentence context was clearly one of increasing intelligibility above that of isolated words. Listeners of deaf speech achieved higher scores for sentences than for isolated words (Hudgins, 1949; Thomas, 1963; Brannon, 1964; Subtelny, 1977). For experienced listeners, the difference between words and sentences ranged from 29% (Hudgins, 1949) to 49% (Thomas, 1963). For inexperienced listeners, the difference between words and sentences ranged from 9% (Brannon, 1964) to 22% (Thomas, 1963). The advantage of sentence context above isolated words for speech-in-noise testing was also about 20% (Miller et al., 1951; O'Neill, 1957).

A significant difference in studies on the speech of the deaf and those in the speech perception literature is found in the definition of an isolated word. In studies on the speech of the deaf, the test words were produced and heard in isolation. Also, the isolated word was different

TABLE 4

SUMMARY OF SCORES FOR ISOLATED WORDS, SEGMENTED WORDS, AND SENTENCES REPORTED FOR DEAF AND NORMAL SPEAKERS

	Isolated Words		Segmented Words		Sentences	
	Score	Listeners	Score	Listeners	Score	Listeners
Studies on Deaf Speech						
Hudgins, 1949.	42%	Trained			71%	Trained
Thomas, 1963	37%	Experienced			77%	Experienced
	35%	Semi-Naive			70%	Semi-Naive
	28%	Naive			50%	Naive
Brannon, 1964.	17%	Naive			26%	Naive
Subtelny, 1977	35%	Experienced			66%	Experienced
Studies on Normal Speech						
Miller et al., 1951. .			70%*	Semi-Naive	94%*	Semi-Naive
			80%**	Semi-Naive	98%**	Semi-Naive
O'Neill, 1957.			70%*	Naive	88%*	Naive
Lieberman, 1963. . . .			24%+	Naive	100%+	Naive
			59%++	Naive	100%++	Naive

* +12dB signal to noise ratio
** +18dB signal to noise ratio
\+ Redundant sentences
++ Non-redundant sentences

from the words in sentence context (e.g. PB monosyllables versus sentences of "everyday American speech"). In the non-deaf literature, the isolated words were originally produced in context, segmented from the sentence, and presented to the listener in isolation. Thus, the test words were identical in segmented and sentence conditions.

To sum, differences between scores for words and sentences reported in the speech of the deaf, are confounded by several factors. Briefly, these are: (1) differences in vocabulary, (2) differences in production; that is, a word is produced differently in sentences and in isolation, and (3) differences in the degree of sentence context.

Other Factors Which May Influence the Intelligibility Scores of Experienced and Inexperienced Listeners

In addition to the effects of listener experience and context, several other factors were of interest. These included: (1) the relative intelligibility of the test material, (2) the length of the sentence, and (3) the position of the test word in the sentence. There is little direct research on the effects of these factors in the speech of the deaf.

As noted in the Introduction, the speech of the deaf is characterized by numerous segmental and suprasegmental errors. Because of these, a test word which might be relatively intelligible for a normal speaker, whether produced in sentence context or in isolation, would not necessarily receive a similarly high score by a deaf speaker. That is, a word produced by a deaf speaker, containing certain phonemes, may receive a relatively low intelligibility score, not only because of differences in articulatory environment (sentences versus isolation) or contextual cues, but also because the deaf speaker is known to have success producing some phonemes more than others. In the research previously reviewed, the relative intelligibility of the speech stimuli used in obtaining measurements of deaf speech has not been carefully controlled.

Another factor of interest, somewhat related to the effect of context, is the effect of overall sentence length. Specifically, the longer the sentence, the more information (i.e. context) is conveyed to the listener, and hence, the greater the intelligibility score. In the previously reviewed literature, the sentences produced by the deaf children varied in length and complexity. These ranged from sentences produced spontaneously by deaf children in response to picture stimuli (John and Howarth, 1965; Markides, 1970), to different sets of the CID sentences of "everyday American speech" (Thomas, 1963; Brannon, 1964). Only Brannon (1964) reported that as the number of syllables in the sentence increased, the overall intelligibility increased, but the differences between sentences varying in length were not large.

Finally, for normal speakers, a test word is known to be produced differently in a sentence depending on its position in the utterance. Specifically, test words produced in the final position of utterances tend to be longer than non-final words (Oller, 1973; Klatt, 1975). However, Nickerson et al. (1974) observed that when there was evidence of pre-pausal lengthening in deaf speech, the increase in the duration of the final syllable was much smaller, on the average, for deaf speakers than for normal hearing speakers. While the effect of word position has been studied relative to use of durational cues in normal or deaf speech, its interest in this study is only with respect to overall intelligibility. That is, are words produced near the end of the sentence more intelligible than words near the beginning or middle of the sentence?

Conclusions

In sum, the review of the literature concludes that there is a difference between experienced and inexperienced listeners in understanding the speech of the deaf and that for both groups of listeners, there is a difference in test scores between sentences and isolated words. Also, several studies have shown that when the listener is able to watch the deaf speaker, intelligibility scores are higher. However, the investigations do not provide information on the reasons for the differences between listeners. This absence of information suggests several questions which may reveal important information to quantify the differences between experienced and inexperienced listeners.

1. Are the differences in test scores between words and sentences the result of differences in production? That is, a word may be produced differently in isolation and in sentence context.

However, in the investigations on the speech of the deaf, the words produced in context were not the same words produced in isolation. Hence, differences in production are also confounded by differences in vocabulary, and no direct answer to the above question can be derived from the literature review.

2. Are the differences in test scores between words and sentences also the result of the semantic context which effects a word in a sentence but not in isolation? Since the sentence material varied considerably from investigation to investigation, and the degree of context for a given set of sentences was not specified, the effect of context on experienced and inexperienced listeners' scores cannot be quantified from the literature.

3. What is the effect of the relative intelligibility of the words on the test scores obtained by experienced or inexperienced listeners? None of the investigations reviewed have attempted to control for the relative intelligibility of a word when produced by a deaf child, thus its influence on test scores obtained by the experienced and inexperienced listeners is unknown.

4. What is the effect of sentence length on the test scores obtained by experienced and inexperienced lis-

teners? In other words, does intelligibility increase as the overall sentence length increases? As previously noted in the investigation on the speech of the deaf, the sentence material varied greatly, and there was little effort to control for this variable of sentence length. Hence, the effect of sentence length on intelligibility scores cannot be answered from a review of the literature.

5. Does production of a word in different positions of the sentence affect intelligibility? None of the investigations reviewed have examined this variable.

While the literature has also shown that training effects and visual information may contribute to differences between listeners, an investigation of these variables is beyond the scope of this study. Thus, the differences between experienced and inexperienced listeners will be investigated by quantifying the effects of those questions posed above, on intelligibility scores obtained in the acoustic domain.

CHAPTER III

EXPERIMENTAL PROCEDURES

Listeners

Listeners to Deaf Speech

One of the key purposes of the study was to quantify the difference between experienced and inexperienced listeners in understanding deaf speech. In order to accomplish this goal, large groups of experienced and inexperienced listeners served as subjects. Sixty listeners were highly experienced in listening to the speech of the deaf; sixty listeners had no previous experience in hearing deaf speech.

An experienced listener was defined as one having had more than one year's experience in listening to the speech of the deaf. The 60 experienced listeners were teachers of the deaf, speech pathologists, and audiologists who were recruited from schools for the deaf. None of the experienced listeners knew the child whose speech they heard or the school at which the child received his training. The number of years experience ranged from just over 1 year to 25 years. The mean number of years experience was 6.8 years.

The 60 experienced listeners ranged in age from 23 to 56 years with a mean age of 28 years. In addition to meeting the experience criterion, each of the listeners was also required to have normal hearing and to be a native speaker of English.

There were 60 inexperienced listeners. An inexperienced listener was defined as one having no previous experience in hearing the speech of the deaf. The inexperienced listeners were recruited among relatives, friends, and undergraduate students in nearby college programs. These listeners were also required to have normal hearing and to be native English speakers.

The 60 inexperienced listeners ranged in age from 18 to 64 years, with a mean age of 27.4 years.

Listeners to Hearing Children's Speech

Since the speech stimuli produced by the deaf children were manipulated in certain listening conditions (described in detail later in this chapter), a group of normal speaking, hearing children served as a control. These speech production tapes were heard by an entirely different group of listeners than those previously described.

A total of 24 listeners recruited primarily from undergraduate classes formed this group of listeners. These listeners may be considered inexperienced in the sense that they had not previously participated in a study to evaluate either deaf or normal speech, and they did not receive any training prior to listening for this study.

The criteria for their selection was as follows:

Each listener had normal hearing and was a native speaker of English.

The 24 listeners ranged in age from 19-56 years, with a mean age of 25.4 years.

Subjects

Deaf Subjects

The deaf subjects were 20 deaf children who were enrolled in the Lexington School for the Deaf. This day and residential school adheres to an oral philosophy of deaf education. Speech training and personal amplification is an integral part of the education program.

The children were equally divided in age groups of 8 to 10 (at least 8 years old and less than 11 years) and 13 to 15 (at least 13 years old and less than 16 years). There were 5 males and 5 females in each age group. Eight years was designated the youngest age since the children were required to read sentence material for collection of a speech sample. The two age groups assured differences in pitch with respect to voice changes due to puberty. These age groups were also selected since there already exists a large body of data on the speech production skills of similar age deaf children for purposes of comparison (Smith, 1972; Levitt et al., 1976).

The subjects were selected from the school register with respect to the following criteria:

1. Congenital Deafness. This was necessary to assure that all subjects had little exposure to speech and language prior to hearing loss.

2. No known physical or psychological anomaly other than deafness. This criterion avoided those children who because of neurological impairment, known brain injury or slow mental development may have additional handicaps which would interfere with the development of speech and language.

3. Enrollment in the selected school for a minimum of two years preceding the study. This assured continuity in the educational and speech training programs.

4. Representative range of speech intelligibility among deaf speakers. This criterion was necessary in order to examine the effect of relative intelligibility for experienced and inexperienced listeners of deaf speech.

The relative intelligibility of the children's speech was determined by conferences with the speech supervisors. No child whose speech was judged to be totally unintelligible was included, since a pilot study, using a different group of subjects, determined that children with unintelligible speech are perceived the same way by experienced and inexperienced listeners. The 20 deaf children in the present study had speech which was judged to be good, average, or fair deaf speech according to the speech supervisors.

For all deaf children, the mean pure tone average of 500 Hz, 1000 Hz, and 2000 Hz was 98.6 dB in the better ear.

For the older deaf children, the mean pure tone average was 98.9 dB in the better ear; for the younger deaf children, it was 97.8 dB in the better ear. Table 5 shows the group mean pure tone averages and the range of the pure tone averages for the deaf children in the study. Appendix A shows the pure tone averages for each of the deaf children in the study.

Hearing Subjects

In addition to the 20 deaf children, 8 hearing children were also tested as a control group. These children were enrolled in public and private school programs in the greater New York area. The children were divided into the same age groups as the deaf children: 8 to 10 years and 13 to 15 years. There were 4 children in the young age group, 2 males and 2 females. There were 4 children in the older age group, 2 males and 2 females.

The children were selected with respect to the following criteria:

1. Normal hearing

2. No physical or psychological anomaly which may have interfered with the normal development of speech and language.

3. Speech production skills considered to be within normal limits as evaluated by a trained speech pathologist.

TABLE 5

MEAN PURE TONE AVERAGE FOR
DEAF CHILDREN GROUPED
FOR AGE AND SEX

	Young Males	Young Females	Old Males	Old Females
Number	5	5	5	5
Mean PTA *	97.8dB	97.8dB	98dB	99.8dB
Range of PTA	78-107dB	82-110+dB	92-105dB	93-107dB

* The mean pure tone average is for the better ear.

Materials

A Priori Estimate of Intelligibility

As noted in the review section, it was necessary to control for relative intelligibility of the test material. However, traditional speech tests were generally not appropriate for this evaluation as the vocabulary was not within the age norms for deaf children in this study. More importantly, detailed intelligibility data were not available on the production of the individual words by deaf children. Test words were therefore selected from data collected by Smith (1972), whose test material was within the vocabulary of deaf children of this age group. All words produced by Smith's deaf children were ranked for intelligibility based on the responses of a large group of inexperienced listeners who transcribed the sentences on a word-by-word basis. These rank scores were examined and a set of 36 monosyllabic words were chosen to form the nucleus of test materials in this study.

The monosyllabic words were as follows:

1. Half of the words were the 18 monosyllabic words which were ranked highest for intelligibility.

2. Half of the words were the 18 monosyllabic words which were ranked lowest for intelligibility.

Table 6 shows the 36 monosyllabic words. Each of the test words was subsequently produced by the deaf and hearing speakers in sentence context and in isolation.

TABLE 6

MEAN SCORES FOR HIGH AND LOW INTELLIGIBILITY MONOSYLLABIC WORDS (SMITH, 1972)

High Intelligibility		Low Intelligibility	
Test Word	Mean Score	Test Word	Mean Score
more	.729	cat	.167
man	.625	food	.167
good	.583	has	.167
this	.562	keep	.167
fat	.542	could	.146
book	.437	feed	.146
fall	.437	that	.125
read	.375	ball	.119
coat	.375	lid	.119
will	.333	teeth	.104
piece	.333	was	.104
have	.292	deep	.104
name	.292	hair	.083
cake	.292	wish	.062
red	.292	beach	.042
need	.262	dog	.042
hear	.250	cool	.042
with	.250	his	.021

A Priori Estimate of Context

In order to examine the effects of context on word intelligibility each of the 36 test words was embedded in a sentence. The sentences contained either a high or low amount of contextual information. In addition, the vocabulary in the sentences was within the range for deaf children 8 to 15 years (Silverman-Dresner and Guilfoyle, 1972).

A definition of high or low contextual information was established for the 36 sentences using a standard word prediction technique. Ten undergraduates read each of the sentences. The test word was omitted from the copy and the students were asked to fill-in the word most appropriate to complete the sentence. A sentence was defined as having high contextual information if the majority of the undergraduates completed it using the deleted word. A sentence was defined as having low contextual information if the undergraduates selected different words to complete the sentence. The design of the 36 sentences was as follows:

1. Eighteen sentences contained a high amount of contextual information. In 9 of these sentences, the underlined test word in Table 7, set A received a high intelligibility ranking. In the other 9 sentences, the underlined test words in Table 7, set B received a low intelligibility ranking.[1]

2. Eighteen sentences contained a low amount of contextual information. These sentences were also balanced for *a priori* intelligibility of the monosyllabic test words.

Table 8, set C, shows the 9 underlined test words with high intelligibility ranking; Table 8, set D, shows the 9 underlined test words with low intelligibility ranking.[2]

Thus, the interaction between context (high versus low) could be systematically investigated. For example, those test words which received a high intelligibility ranking and were in sentences with high context (set A) were expected to receive the highest scores. These sentences provided the listener with the greatest possible combination of information for all variables in the sentence condition.

On the other hand, the test wordswith low intelligibility ranking in sentences with low context (set D) provided the listeners with a paucity of information and were therefore expected to receive the lowest overall scores.

These two sets of stimuli with high context-high intelligibility and low context-low intelligibility represent opposing examples of the interaction between sentence context and the intelligibility of the test word. In two additional sets of sentences (sets B and C), the intelligibility of the test word and context were systematically varied. That is, sentences with high context had test words with low intelligibility; sentences with low context had test words with high intelligibility. In these sentences, the central interest was to determine if the a priori estimate of intelligibility of the test word could be changed as a consequence of sentence context.

TABLE 7

SENTENCES WITH HIGH CONTEXT CONTAINING TEST WORDS WITH HIGH OR LOW A PRIORI INTELLIGIBILITY

High Context, High Intelligibility (Set A)	High Context, Low Intelligibility (Set B)
3 Syllables	3 Syllables
Read the book. *	Keep quiet. *
Come with me.	The dog barks.
That's no good. *	Comb your hair.
5 Syllables	5 Syllables
My name is Nancy.	The cat chased the mouse.
Get your coat and hat.	Get your ball and bat.
Is there no more milk?	Did you brush your teeth?
7 Syllables	7 Syllables
That man is not my father.	I wish I had a pony.
The flag is red, white and blue.	We have food for the picnic.
May I have a piece of cake?	Can you dive in deep water?

* These sentences were defined as lower in context than others in the table.

TABLE 8

SENTENCES WITH LOW CONTEXT CONTAINING TEST WORDS WITH HIGH OR LOW A PRIORI INTELLIGIBILITY

Low Context, High Intelligibility (Set C)	Low Context, Low Intelligibility (Set D)
3 Syllables	3 Syllables
Have a lot.	Feed the dog.
I need it.	You did it.
Get the cake.	This is his.
5 Syllables	5 Syllables
They will come again.*	Is that the tall one?
Who wants this ice cream?	Mother has the car.
It's easy to hear her. +	He said he could go.
7 Syllables	7 Syllables
The book is on the table.	What was the name of that boy?
Is the fat baby crying?	If it's cool I cannot go.
It is nice on a fall day.	We will go to the beach today. +

* This sentence was defined as higher in context than others in the table.
\+ These sentences contain an additional syllable.

Number of Syllables and Word Position in the Sentence

The 36 sentences have also been developed with respect to other factors which might result in differences in intelligibility. These include: (1) the number of syllables in the sentence, and (2) the location of the test word in the sentence.

The 36 sentences were arranged with respect to length as follows: 12 were three syllables in length, 12 were five syllables in length, and 12 were seven syllables in length.[3] The 3, 5 and 7 syllable sentences contained an equal number of high and low intelligibility test words, and were equally divided with respect to high and low sentence context. Thus, in each set of 9 sentences with high intelligibility-high context, 9 with low intelligibility-high context, etc., there were an equal number of 3, 5 and 7 syllable sentences. This design is apparent in the grouping of sentences arranged in Tables 7 and 8.

The location of the test word in the sentence was also varied. For 3 syllable sentences, the test word occurred either: (1) in the word initial position; (2) in the middle of the sentence, or (3) in the word final position. For 5 and 7 syllable sentences, the test word occurred either: (1) near the beginning of the sentence, (2) in the middle of the sentence, or (3) near the end of the sentence. Each of the test words is underlined in Tables 7 and 8, and the location of the test word in the sentence is fairly obvious.

Summary of the Design of the Test Materials

To sum, the factors of interest in the test materials are: (1) *a priori* intelligibility of the test word, (2) sentence context, (3) the number of syllables in the sentence, and (4) the position of the test word in the sentence. These are schematically diagrammed in Figure 1.

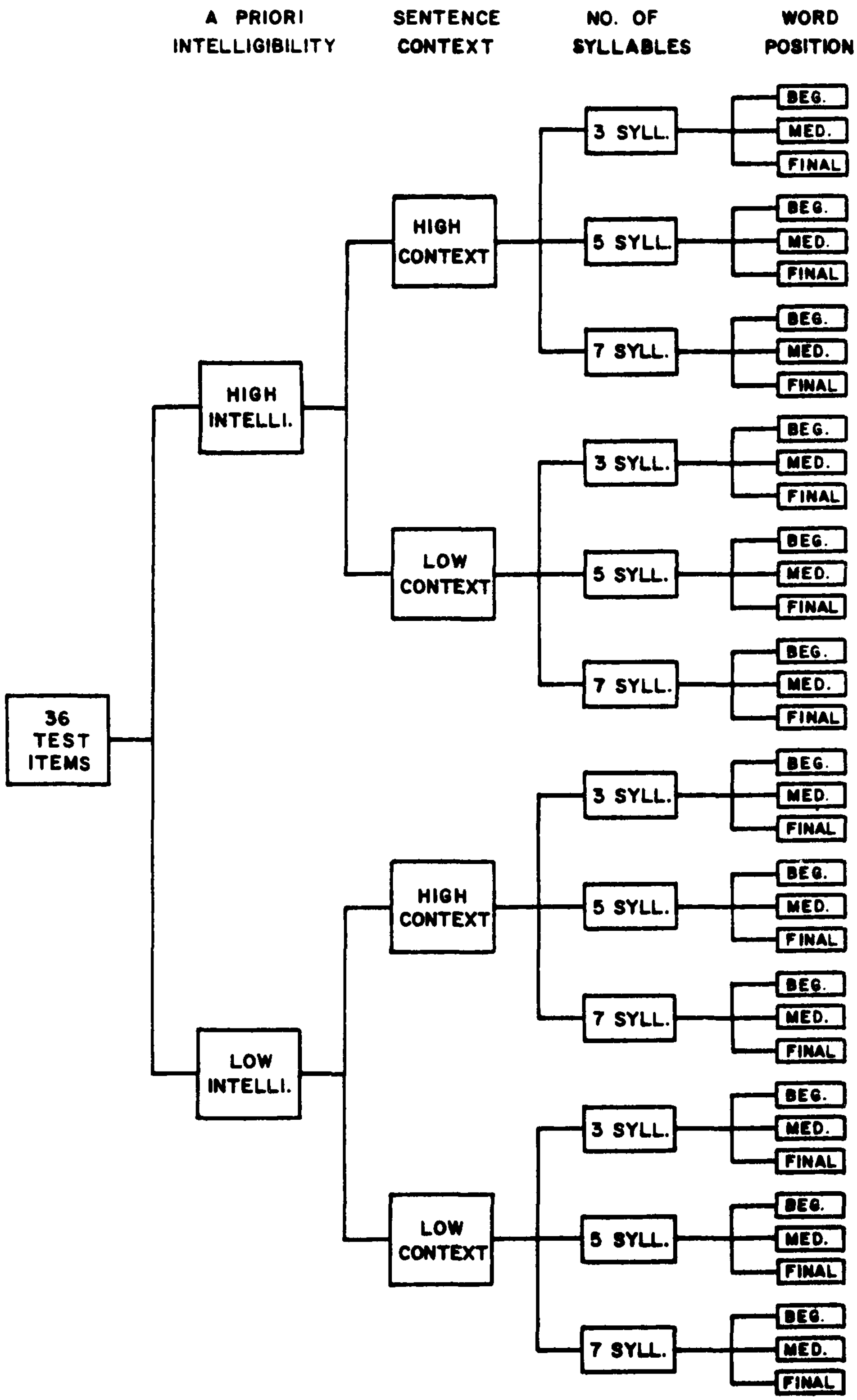
For the 36 test words in sentences, all factors through word position are relevant to this type of speech stimulus. Since the 36 segmented test words (to be described in the following sections of this chapter) were originally produced in sentences, all factors are also applicable to this type of speech stimulus. For the 36 test words in isolation, only *a priori* intelligibility is a factor.

Recordings of the Children's Speech

The speech samples collected from the deaf and hearing children were recorded on a Tandberg Model 11 tape recorder using low print-through tape (Scotch #206) and a Norelco microphone, Model D24E.

The recordings of the deaf children's speech were made in a sound proof booth with the child positioned to face the microphone at a distance of approximately 10 inches from the lips. Each child was seen for a recording session approximately 20 minutes in length. No practice preceded the recording of the speech stimuli. Each deaf child read the list of words and sentences twice, since Smith (1972) demonstrated considerable variability between practice and

Figure 1. A schematic diagram of the design of the test materials.


A PRIORI INTELLIGIBILITY
SENTENCE CONTEXT
NO. OF SYLLABLES
WORD POSITION
36 TEST ITEMS
HIGH INTELLI.
LOW INTELLI.
HIGH CONTEXT
LOW CONTEXT
HIGH CONTEXT
LOW CONTEXT
3 SYLL.
5 SYLL.
7 SYLL.
BEG.
MED.
FINAL

performance recording by deaf children. The production of the word or sentence which seemed more intelligible to the experimenter was selected for later analysis.

Recordings of the young hearing children's speech were made in relatively quiet areas in their school. The speech samples of 3 of the older hearing children were recorded in a sound proof booth; the fourth older hearing child was recorded in a relatively quiet area of his home. The children were positioned to face the microphone at a distance of approximately 10 inches from the lips. Each recording session for the hearing children was approximately 10-15 minutes in length. The hearing children read each word or sentence once. In the rare event that an error occurred, the child simply re-read the item.

Listening Conditions

The speech recording of the deaf and hearing children were prepared in order to present three types of speech stimuli to the listeners. These were:

1. Test words produced in sentences and presented to the listeners in sentences.

2. Test words produced in isolation and presented to the listeners in isolation.

3. Test words produced in sentences, segmented from the sentences, and presented to the listeners in isolation. Hereafter, these are referred to as segmented test words.

In order to obtain the segmented test words for the third listening condition, the audio tape-recordings of the sentences described in the previous section, were processed using the Ubiquitous Spectrum Analyzer (Federal Scientific UA-6A) and the Honeywell DDP 224 computer system at Haskins Laboratories, New Haven. The system was used only as an editing device to segment the test words from sentence context.

The sentences were entered into the computer system using the Ubiquitous Spectrum Analyzer and an input-editing program (SEDIT). This program was used to create disc files of edited waveform or spectrum data. The waveform data, sampled at 10 KHz, was bandpassed at 90 Hz-3.2 Hz. Spectrum data was the output of the Ubiquitous Spectrum Analyzer, typically 128 samples per frequency scan (approximately 4900 Hz bandwidth), repeated every 12.8 msecs.[4]

The speech was stored in digital form by inputting the waveform and the spectrum data to magnetic tape (in real time), and then copying the data to an input file on disc. Waveform playback and spectrogram displays could then be used to edit the test words from each sentence.

The waveform display (up to 1.6 sec) was viewed on an oscilloscope screen. The SEDIT program permitted two pointers to be independently moved across the waveform display in 12.8 msec steps. The pointers were repeatedly adjusted until the test word in the sentence was isolated. It was also possible to play the portion of the acoustic waveform between the pointers. Thus, by viewing the waveform on the oscilloscope and by listening to the segmented portion, it was possible to isolate each test word from the sentence.

The waveform and spectrum data for the segmented test word were temporarily stored on a disc file. The 36 segmented test words were recorded when the entire segmentation process was completed for an individual child.

One limitation of the system was that the pointers used to segment the test words from the sentences could not be continuously adjusted. However, segmenting a test word from a sentence produced by a deaf child proved to be relatively easy in most cases, since many of the deaf children produced the sentences on a word-by-word basis with relatively large pauses between words. Figure 2 shows the waveform for an older deaf child's production of the sentences "Read the book." with the pointers A and B indicating the segmented test word "read." In the rare cases for which there were no clear cut pauses between words in the sentences, as in Figure 3, spectrograms were also made of the edited test word, and the entire sentence which served to confirm the segmentation procedure.

Spectrograms were made using the Haskins' CONVERT program.[5] Figure 4 shows the spectrogram for the sentence and the segmented test word whose waveform has been illustrated in Figure 3.

The same procedure was followed for segmenting the test words from the hearing children's sentences. In some instances, segmenting the test words from the hearing chil-

Figure 2. The waveform of a sentence produced by a deaf child with the pointers indicating the test word for segmentation.

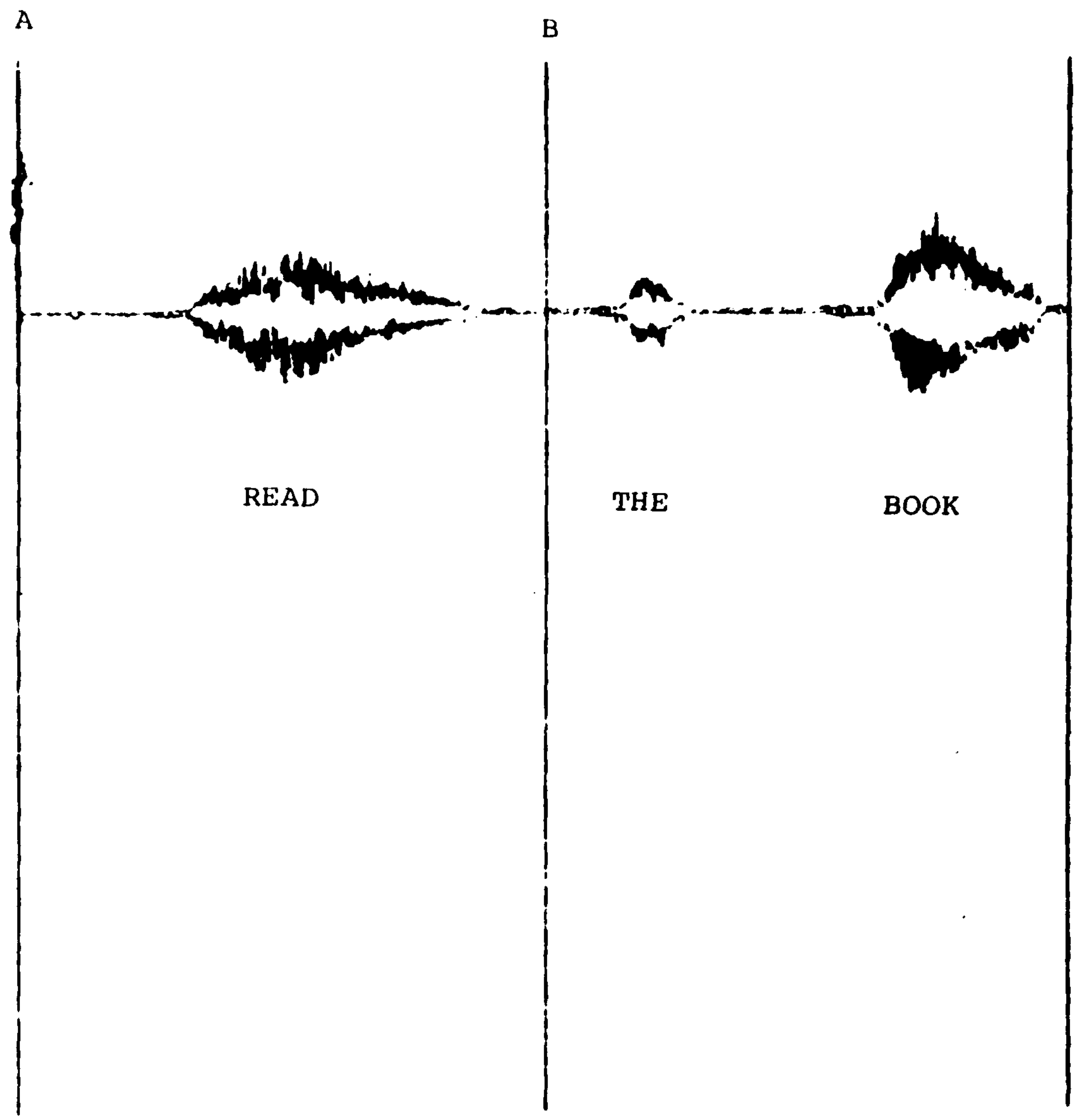

A
B
READ
THE
BOOK

Figure 3. The waveform of a sentence produced by a deaf child for which segmentation of the test word was difficult.

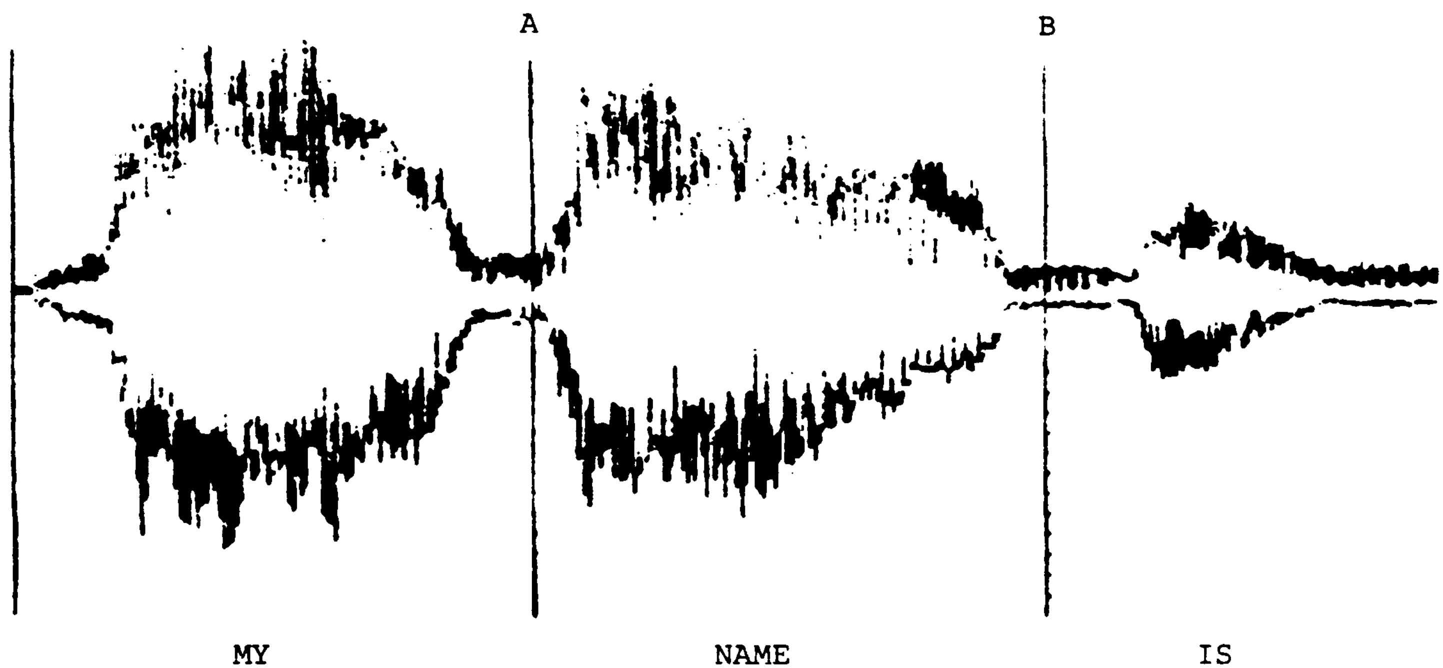
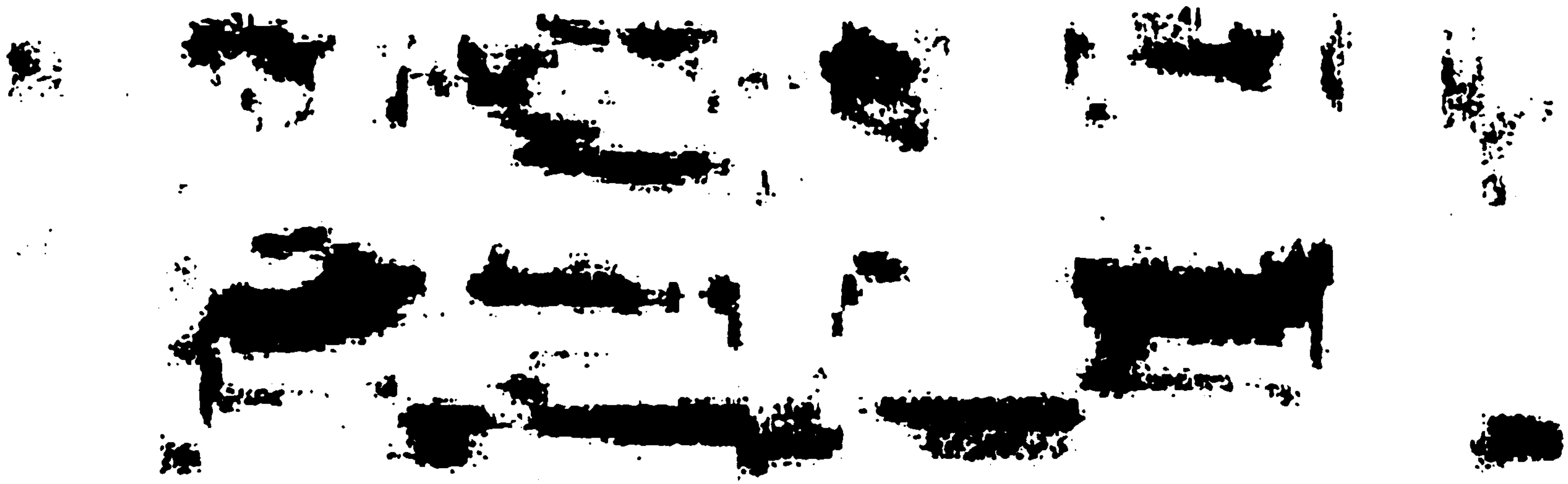
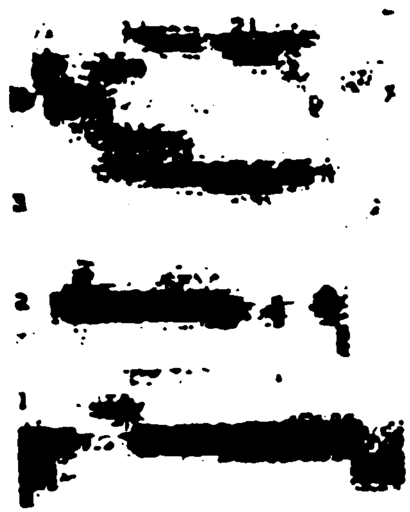

A
B
MY
NAME
IS

Figure 4. Spectrogram of the sentence and test word for the waveform illustrated in Figure 3.



MY NAME IS NAN . . . . . CY



NAME

Figure 5. The waveform of a sentence produced by a hearing child with the pointers indicating the test word for segmentation.

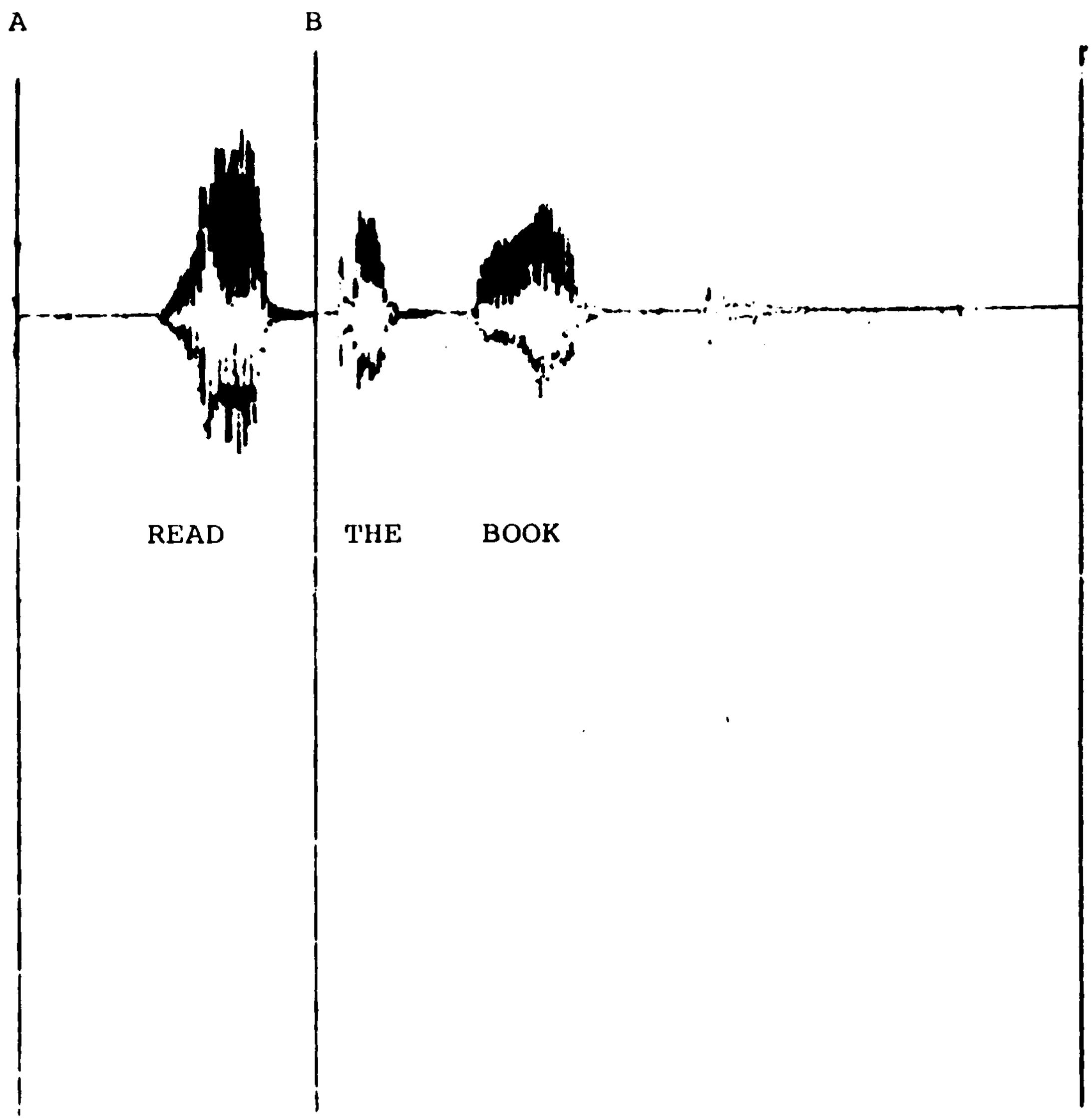

A
B
READ
THE
BOOK

Figure 6. Spectrogram of the sentence and test word for the waveform illustrated in Figure 5.

READ THE BOOK

dren's sentences proved slightly more difficult than for the deaf children because of coarticulation effects. However, using the combined visual and auditory procedures described, word boundaries could be determined.

Figure 5 shows the waveform for an older hearing child's production of the sentence "Read the book." with pointers A and B indicating the segmented test word "read." Figure 6 shows the spectrogram for the same child's production of the sentence.

Specification of Speech Levels

During input to the computer system, the overall intensity level for each of the sentences was frequently changed in order to make maximum use of the available dynamic range of the digital storage. Thus, the intensity level of the segmented word was frequently quite different than the intensity level of the same word in the sentence. Similarly, production of the test word in isolation also resulted in intensity levels different from those for the same word in a sentence or segmented. In summary then, the intensity levels of the same word produced in 3 different conditions (sentence, isolated or segmented) were different either because of production differences or because of manipulation. These differences in intensity alone might account for differences in intelligibility, and thereby confound analysis of the effects of articulatory environment, a priori intelligibility, sentence context, etc.

In order to minimize this problem, level recordings were made of the speech stimuli using the Bruel and Kjaer Level Recorder (Model 2305). Levels were measured for each of the test words produced in sentences, in isolation, and segmented. The intensity level (in decibels) of each test word in the sentence condition was set as the reference level. Then, the levels for both the test word in isolation and the segmented test word were adjusted to within $\pm$ 1 dB of the reference level. These stimuli were re-recorded. Thus, the test word in all three conditions was presented at a common intensity level within $\pm$ 1 dB. Appendix B shows the block diagram of the equipment used to equalize the speech stimuli and the description of the equalizing procedure.

Following the equalizing procedure, a 400 Hz calibration tone was generated which preceded each child's 3 sub-master tapes: test words in sentences, test words in isolation, and segmented test words. The level of the calibration tone was the mean of the 36 test words in sentences for each child. This calibration tone was used to maintain the equalized levels during the randomization procedure and during playback.

Preparation of the Listening Tapes

The set of equalized sub-master tapes were randomized, re-recorded and distributed on 84 listening tapes. A tape contained 36 items produced by a single deaf or a single hearing child. These items were the various productions (sentences, isolated or segmented test words) of the 36 test words described previously in this chapter. In order to avoid later learning effects by the listeners, there was no repetition of a test word in any condition on a tape. A child's total production of speech stimuli was therefore distributed over 3 tapes. Sixty tapes were produced for the deaf children (3 tapes times 20 children); 24 tapes were produced for the hearing children (3 tapes per child times 8 children).

Each of the production conditions were arranged on a tape as follows:

1. Twelve items were test words produced in sentences and presented to the listeners in sentences;

2. Twelve items were test words produced in isolation and presented to the listeners in isolation;

3. Twelve items were test words produced in sentences and presented to the listener in isolation (i.e. segmented).

A strip of leader tape separated the 3 aforementioned conditions on each tape. Each of the 12 items was selected for a tape in such a manner that a tape contained an equal number of test words with high or low a priori intelligibility. Also, sentence context (high or low), number of syllables (3,5 or 7) and position of the test word in the sentence (near the beginning, middle or end) were also distributed across the test items as appropriate (refer to Figure 1). Appendix C shows the sequence of the test items and the relevant factors for each on tapes 1, 2, and 3.

A perusal of the items on each tape shown in Appendix C will reveal that the order of the conditions varied on each tape. On the first tape produced for each child, the first 12 items were test words in sentences, followed by segmented test words, and then test words in isolation. On the second tape for each child, the first 12 items were a different set of segmented test words, followed by a different set of test words, in isolation, and finally test words in sentences. On tape 3, the order for the last set of words was: test words in isolation, test words in sentences and segmented test words. Within each group of 12 items the order of the test words were also randomized.

Listening Procedures

In order to avoid learning effects by the listeners, each tape produced by a deaf speaker was heard by one experienced and one inexperienced listener. There were a total of 120 listeners in all--60 experienced and 60 inexperienced listeners. Each of the listeners read a set of typed instructions shown in Appendix D. The listener's task was to write down what he heard. Thus, in the sentence condition, information was also obtained on words other than test words. The listener was told in advance whether he would be hearing a set of words or sentences.

Listening was done monaurally via TDH-49 headphones and a Uher tape-playback system. A 400 Hz calibration tone preceded the tape and allowed the listener to adjust the volume to a comfortable listening level. Once this level was set, it was not changed. The tape was then run by the examiner, who replayed each item to the listener. A listening session lasted approximately 20 minutes.

The listening procedure for the tapes of hearing children was essentially the same. Slightly different instructions were used. These are also shown in Appendix D. A listening session lasted approximately 10 minutes.

The responses were scored similarly for all listeners. For test words in isolation and for segmented test words, the responses were coded as follows: correct, incorrect, or no response. For the main body of the analysis, responses were analyzed only as correct or incorrect; "no response" items were coded as errors.

The responses for all words in sentences were scored on a syllable-by-syllable basis since many of the non-test words in the sentences were bi-syllabic (e.g. Mother has the car.). Although not of central interest in this study, the responses for all words in the sentences were coded. Of particular interest were the responses for the test words in sentences. The responses for test words were analyzed separately. The coding format for test words and non-test words in sentences was the same as described for test words in isolation and segmented test words.

FOOTNOTES

[1]Formal definition of high or low context was made by the undergraduates after the sentences had, in fact, been recorded by the deaf children. In several instances, the formal definition did not agree with the author's informal estimate of context when the sentences were developed. In Table 7, four sentences which had been developed as containing a high degree of context elicited a variety of responses from the undergraduates. These sentences are marked with an *, and may be considered not as high in the degree of context as the other sentences in this Table. However, the sentences are treated as high context sentences in the data analysis.

[2]In Table 8, one sentence which was developed as low in context, elicited the same deleted word from the undergraduates, indicating a high degree of context. This sentence is marked with an *, and may be considered not as low in context as the other sentences in Table 8. However, the sentence is treated as a low context sentence in the data analysis.

[3]Two sentences were inadvertently included which contained 6 instead of 5 syllables, and 8 instead of 7 syllables. The sentences are marked with a + in Table 8.

[4]G.M. Kuhn, "User's Guide to SEDIT," Haskins Laboratories, 1974 (mimeographed).

[5]G.M. Kuhn, "User's Guide to CONVERT," Haskins Laboratories, 1975 (mimeographed).

CHAPTER IV

RESULTS

Results For Deaf Children

Statistical Analysis

The primary purpose of this study was to investigate the differences between experienced and inexperienced listeners in their understanding of the speech of the deaf. Intelligibility scores were obtained for experienced and inexperienced listeners, and analyses of variance performed to test for significant interactions between listener experience and other factors of interest. A listing of the intelligibility scores obtained by experienced and inexperienced listeners for deaf speakers in each of the experimental conditions is shown in Appendix E.

In order to analyze the interaction between listener experience and other factors, analyses of variance were performed for test words in sentences, in isolation, and in segmented conditions. Separate analyses were performed because the number of factors were not similar for each type of stimulus. The factors considered in the analyses included listener experience, *a priori* intelligibility of the test word, degree of sentence context, and two additional factors pertaining to the speakers. These were:

age of the children (younger versus older), and sex (male versus female). The analyses of variance for test words in sentences and for segmented test words included all five factors. The analysis for isolated words had only four factors since context was not a factor for words produced and heard in isolation.

In order to analyze the differences between the types of stimuli, a fourth analysis of variance was done. In this case the factors were: the type of stimulus (test words in sentences, in isolation, and segmented conditions), listener experience, and a priori intelligibility.

In performing the analyses of variance, data was transformed using the arc sine transformation. This was done so as to stabilize the variability that occurs when estimating a proportion from a finite number of items. A derivation and discussion of the transformation appears in Brownlee (1965). The analysis of variance for test words in sentences, in isolation, and in segmented conditions appear in Tables 9, 10 and 11 respectively. The analysis of variance for the differences between types of speech stimuli is presented in Table 12. Because of the large number of F tests performed in each of these analyses, only those effects with a significance level of .01 or smaller were considered. Effects with a significance level of between .01 and .02 are labelled borderline.

Of particular interest was the effect of listener experience and any interaction between experience and other

TABLE 9

ANALYSIS OF VARIANCE FOR TEST WORDS
PRODUCED AND HEARD IN SENTENCES

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
Between Children					
Age (A)	12.04	1	12.04	11.58	.003 **
Sex (S)	1.07	1	1.07	1.03	.326
A x S	1.10	1	1.10	1.06	.320
Explained	14.21	3	4.73	4.54	.048
Between Children Differences	16.58	16	1.04	----	
Total	30.79	19	1.62	----	
Within Children					
Experience (E).	2.440	1	2.44	20.5	.001 **
A x E.	.006	1	.006	.05	.817
S x E.	.250	1	.250	2.10	.146
A x S x E.	.600	1	.600	5.04	.025
A Priori Intell. (I). .	2.730	1	2.730	22.94	.001 **
A x I.	.010	1	.010	.08	.769
S x I.	.630	1	.630	5.29	.021
A x S x I.	.110	1	.110	.92	.659
E x I.	.003	1	.003	.02	.868
A x E x I.	.090	1	.090	.75	.609
S x E x I.	.006	1	.006	.05	.817
A x S x E x I.	.034	1	.034	.28	.600

TABLE 9-Continued

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
Context (C)	6.38	1	6.380	53.61	.001 **
A x C.	.020	1	.020	.16	.685
S x C.	.720	1	.720	6.05	.014 B
A x S x C. . . .	.070	1	.070	.58	.549
E x C.	.007	1	.007	.05	.804
A x E x C. . . .	.013	1	.013	.10	.740
S x E x C. . . .	.001	1	.001	.00	.924
A x S x E x C. .	.516	1	.516	4.33	.037
I x C.	2.202	1	2.202	18.50	.001 **
A x I x C. . . .	.107	1	.107	.89	.652
S x I x C. . . .	.168	1	.168	1.41	.235
A x S x I x C. .	.112	1	.112	.94	.664
E x I x C. . . .	.029	1	.029	.24	.628
A x E x I x C. .	.069	1	.069	.57	.545
S x E x I x C. .	.000	1	.000	.00	1.000
A x S x E x I x C	.191	1	.191	1.60	.205
Explained.	17.51	28	.62	5.20	.021
Residual Sum of Squares.	13.37	112	.119	----	
Total.	30.88	140	.22	----	

** Significant at < .01 level.
B Significant between .01 and .02 level.

TABLE 10

ANALYSIS OF VARIANCE FOR TEST WORDS
PRODUCED AND HEARD IN ISOLATION

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
Between Children					
Age (A)	3.08	1	3.08	11.84	.003 **
Sex (S)	.10	1	.10	.38	.542
A x S	.00	1	.00	.00	1.000
Explained	3.18	3	1.06	4.07	.04
Between Children Differences	4.29	16	.26	----	----
Total	7.47	19	.39	----	----
Within Children					
Experience (E).	.442	1	.440	14.6	.001 **
A x E.	.006	1	.006	.2	.675
S x E.	.084	1	.080	2.66	.113
A x S x E.	.000	1	.000	.0	1.000
A Priori Intell. (I). .	2.313	1	2.310	70.9	.001 **
A x I.	.002	1	.002	.06	.809
S x I.	.007	1	.007	.23	.651
A x S x I.	.015	1	.015	.45	.510
E x I.	.010	1	.010	.33	.590
A x E x I.	.003	1	.003	.10	.760
S x E x I.	.001	1	.001	.03	.850
A x S x E x I.	.045	1	.045	1.5	.240
Explained	2.92	12	.24	8.0	
Residual Sum of Square	1.63	48	.03	----	----
Total	4.55	60	.07	----	----

** Significant at ∠.01 level

TABLE 11

ANALYSIS OF VARIANCE FOR TEST WORDS PRODUCED IN SENTENCES AND HEARD IN ISOLATION (SEGMENTED)

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
Between Children					
Age (A)	.706	1	.706	4.46	.048
Sex (S)	1.206	1	1.206	7.63	.013 B
A x S	.147	1	.147	.93	.648
Explained	2.05	3	.683	4.32	.04
Between Children Differences	2.54	16	.158	----	----
Total	4.59	19	.241	----	----
Within Children					
Experience (E).	.292	1	.292	5.03	.025
A x E.	.023	1	.023	.39	.537
S x E.	.001	1	.001	.01	.891
A x S x E.	.000	1	.000	.0	1.000
A Priori Intell. (I). .	.884	1	.884	15.24	.001 **
A x I.	.006	1	.006	.10	.747
S x I.	.108	1	.108	1.86	.171
A x S x I.	.090	1	.090	1.55	.213
E x I.	.001	1	.001	.01	.891
A x E x I.	.035	1	.035	.60	.554
S x E x I.	.049	1	.049	.84	.637
A x S x E x I.	.052	1	.052	.89	.652

TABLE 11-Continued

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
Context (C)	1.740	1	1.740	30.00	.001 **
A x C.	.020	1	.020	.34	.565
S x C.	.233	1	.233	4.01	.044
A x S x C.	.107	1	.107	1.8	.173
E x C.	.035	1	.035	.60	.554
A x E x C.	.027	1	.027	.46	.503
S x E x C.	.045	1	.045	.77	.615
A x S x E x C.	.025	1	.025	.43	.520
I x C.	.169	1	.169	2.91	.086
A x I x C.	.099	1	.099	1.70	.191
S x I x C.	.031	1	.031	.53	.526
A x S x I x C.	.152	1	.152	2.62	.104
E x I x C.	.342	1	.342	5.89	.015 B
A x E x I x C.	.022	1	.022	.37	.546
S x E x I x C.	.037	1	.037	.63	.568
A x S x E x I x C. .	.001	1	.001	.01	.891
Explained.	3.75	28	.13	2.24	.14
Residual Sum of Squares.	6.51	112	.058	----	----
Total.	10.26	140	2.24	----	----

** Significant at ∠.01 level
B Significant between .01 and .02 levels

TABLE 12

ANALYSIS OF VARIANCE FOR LISTENER EXPERIENCE, TYPE OF STIMULI AND *A PRIORI* INTELLIGIBILITY OF TEST WORD

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
Listener Experience (E) .	5.14	1	5.14	29.11	.000
Type of Stimulus (T) . .	31.54	2	15.77	89.34	.000
A Priori Intelligibility of Test Word (I) . . .	11.72	1	11.72	66.39	.000
E x T.	1.01	2	.50	2.86	.058
E x I.	.08	1	.08	.47	.491
T x I.	1.09	2	.54	3.07	.045
E x T x I.	.04	2	.02	.104	.902
Explained.	50.62	11	4.60	26.07	.000
Residual	760.48	4308	.177	----	----
Total.	811.10	4319	.188	----	----

factors. Tables 9, 10, and 11 show that experience was highly significant for test words in sentences and in isolation, but about the borderline significance level for segmented test words. There was no interaction between experience and any of the other factors for test words in sentences, in isolation, or in segmented conditions.

Other significant main effect included: context, *a priori* intelligibility, and age (the latter factor was significant only for test words in sentences and in isolation). Sex was not a significant factor.

There was some evidence of an interaction between *a priori* intelligibility and context (IxC) for test words in sentences, and also a borderline interactive effect between sex and context (SxC). There was also a borderline interactive effect between experience, *a priori* intelligibility, and context (ExIxC) for segmented test words.

The Differences Between Experienced and Inexperienced Listeners

Of primary interest was the difference between experienced and inexperienced listeners. Table 13 summarizes the mean scores obtained by experienced and inexperienced listeners for each type of speech stimulus. Scores for experienced listeners were consistently higher than those for inexperienced listeners. For both groups, scores for test words in sentences were consistently higher than those for isolated words which in turn were higher than the scores for segmented words. Scores for test words in sentences were more than double the scores for segmented words.

TABLE 13

MEAN SCORES OBTAINED BY EXPERIENCED AND INEXPERIENCED LISTENERS

Type of Stimulus	Listeners	Mean Score % Correct	Mean Score Arc Sine Units
Test words produced and heard in sentences.	Experienced	.41	.018
	Inexperienced	.30	.017
Test words produced and heard in isolation.	Experienced	.29	.017
	Inexperienced	.23	.016
Test words produced in sentences and heard in isolation (i.e. segmented).	Experienced	.16	.014
	Inexperienced	.13	.012
All words produced and heard in sentences.	Experienced	.49	.015
	Inexperienced	.35	.014

The greatest difference between listeners occurred on sentences. Scores for test words in sentences were 41% for the experienced and 30% for the inexperienced listeners, a difference of eleven percentage points. In contrast, the difference between listeners for words in isolation was six percentage points, and three percentage points for segmented test words. Also shown in Table 13, are the average arc - sine values.

All significant differences, including all reported interactions between listener experience and other factors, are based on the analysis of variance in which the arc-sine transformation was used. The average scores are reported in percentages (rather than arc-sine units) for the reader's convenience. However, it should be remembered that the arc-sine transformation significantly expands the measurement scale for test scores that are close to 100% or close to 0%. For example, a change in score of five percentage points in the vicinity of 50% (such as from 50% to 55%) is equivalent in the arc-sine domain to a two percent change in scores in the vicinity of a 5% or 95% score (that is, from 5% to 7%, or from 93% to 95%). This effect should be kept in mind when assessing the magnitude of interactive effects as reported in terms of average percentage scores.

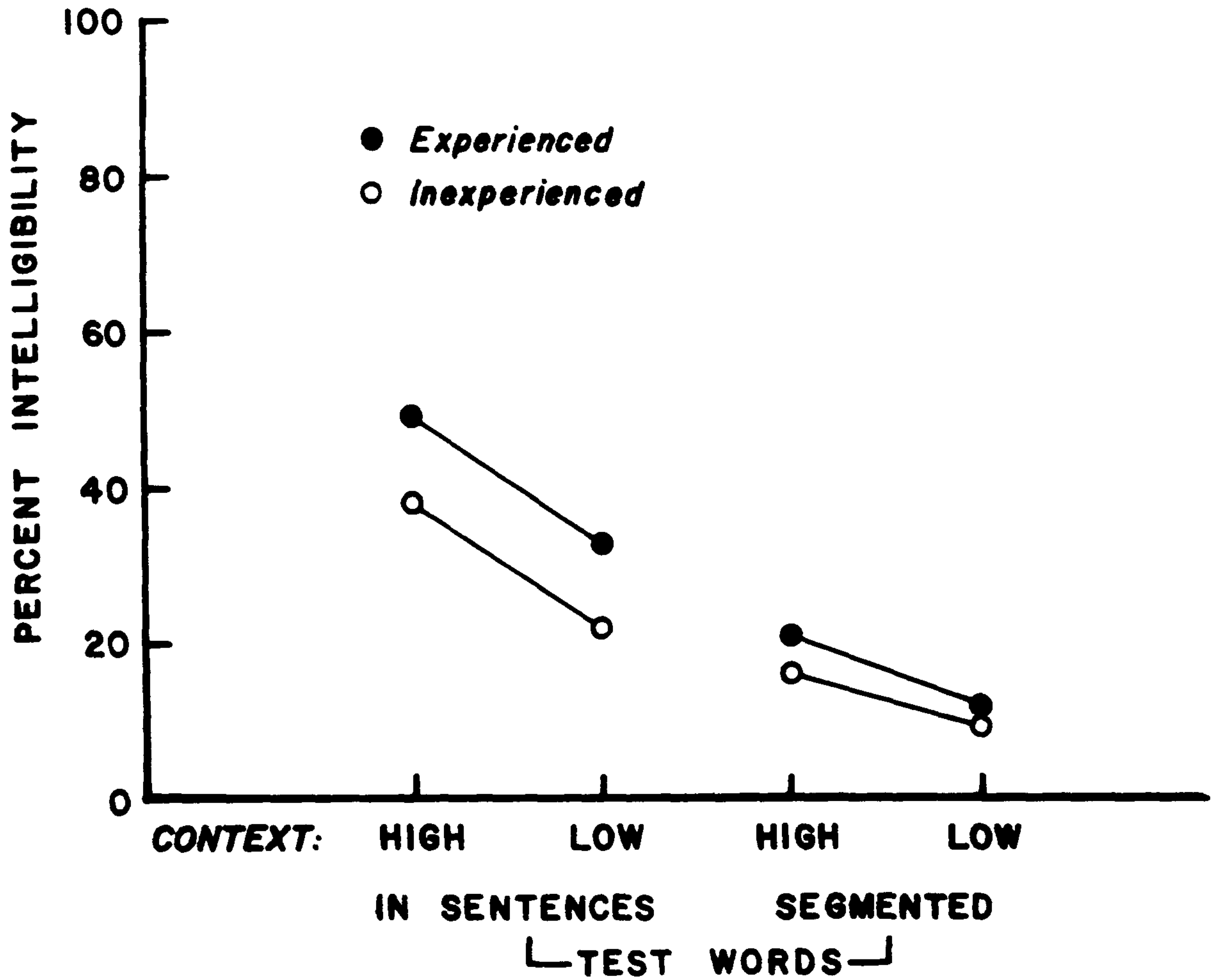
Intelligibility scores were also obtained for all words in sentences. Although not of central interest, these data are also included in Table 13. Scores for all words in sentences were 49% for the experienced and 35% for the inexperienced listeners. The scores for all words were higher by a smaller amount than for scores based on test words only.

Degree of Sentence Context

Scores obtained by experienced and inexperienced listeners for test words as a function of the degree of sentence context are plotted in Figure 7. The percent intelligibility is plotted on the ordinate; the degree of sentence context (high or low) is plotted on the abscissa. The average values in this plot were obtained from the analyses of variance in inverse sine units and were then transformed back to percentages. These values are similar, but not identical to the original scores. This transformation has been used in plotting all data throughout this chapter, unless otherwise noted.

Figure 7 summarizes the data for test words in sentences as well as for segmented test words since the latter were originally produced in sentence context. The left hand side shows data for test words in sentences; the right hand side shows data for segmented test words. On average, the experienced listeners' scores were higher than the inexperienced listeners. There was no evidence of an interaction between experience and context since the improvement in intelligibility due to experience was the same for both high and low context words. This difference was roughly equal to a 10% gain in intelligibility for

Figure 7. Mean scores obtained by experienced and inexperienced listeners for test words as a function of sentence context.


100
80
60
40
20
0
PERCENT INTELLIGIBILITY
Experienced
Inexperienced
CONTEXT:
HIGH
LOW
HIGH
LOW
IN SENTENCES
SEGMENTED
TEST WORDS

sentences, and roughly a 5% gain for segmented words.

The magnitude of the context effect is also evident from Figure 7. Scores for both experienced and inexperienced listeners were greater for the high context sentences than for the low. Scores for test words in high context sentences were 49% for the experienced and 38% for the inexperienced listeners; scores for sentences in low context sentences were 33% for the experienced and 21% for the inexperienced listeners. This represents a gain of roughly 16 percentage points for either experienced or inexperienced listeners.

In contrast, scores for test words segmented from high context sentences were 21% for the experienced listeners and 16% for the inexperienced listeners. Scores for test words segmented from low context sentences were 11% for the experienced listeners and 9% for the inexperienced listeners. This represents an average gain of about 8 percentage points for either group of listeners.

Thus, the effect of context for words produced and heard in sentences is substantial. If the same test words are segmented, such that, although produced in context they are heard in isolation, the effect of context is much smaller, but not negligible.

A Priori Intelligibility

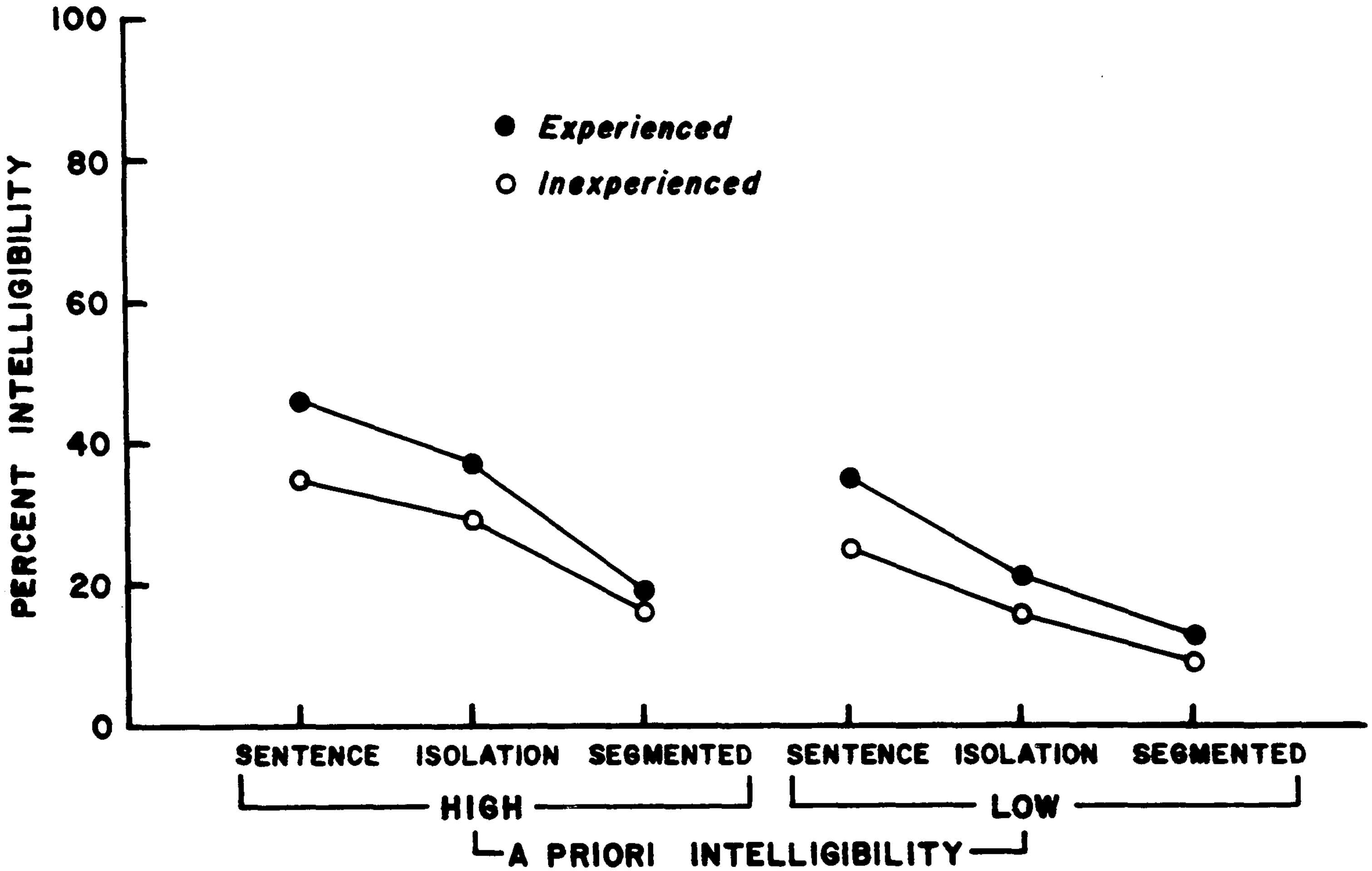
The relative intelligibility of a test word as a function of a priori intelligibility was also examined. Scores obtained by experienced and inexperienced listeners are plotted in Figure 8. The percent intelligibility is plotted on the ordinate; the three types of speech stimuli are plotted on the abscissa. As before, experienced listeners obtained higher scores than inexperienced listeners for either high or low intelligibility words in sentence, in isolation, and in segmented conditions.

Figure 8 shows that the overall pattern of the data for high and low intelligibility words was similar for both groups of listeners. Test words with high *a priori* intelligibility received higher scores than those with low *a priori* intelligibility in each experimental condition. For either high or low intelligibility words, scores were highest when the test words were in sentences, followed by test words in isolation, and finally segmented test words. However, the effect of *a priori* intelligibility was most pronounced for test words in sentences and in isolation. In these conditions, scores obtained by experienced and inexperienced listeners were noticably higher for test words with high *a priori* intelligibility than with low. High or low *a priori* intelligibility had less effect on the scores for segmented words. The interaction between *a priori* intelligibility and stimulus type reached the .05 level of significance.

Interaction Between Experience, Context, and Intelligibility

Of special interest was the significant interaction between intelligibility and context (IxC) for sentences as well as any interaction involving experience and these

Figure 8. Mean scores obtained by experienced and inexperienced listeners for test words as a function of _a priori_ intelligibility.

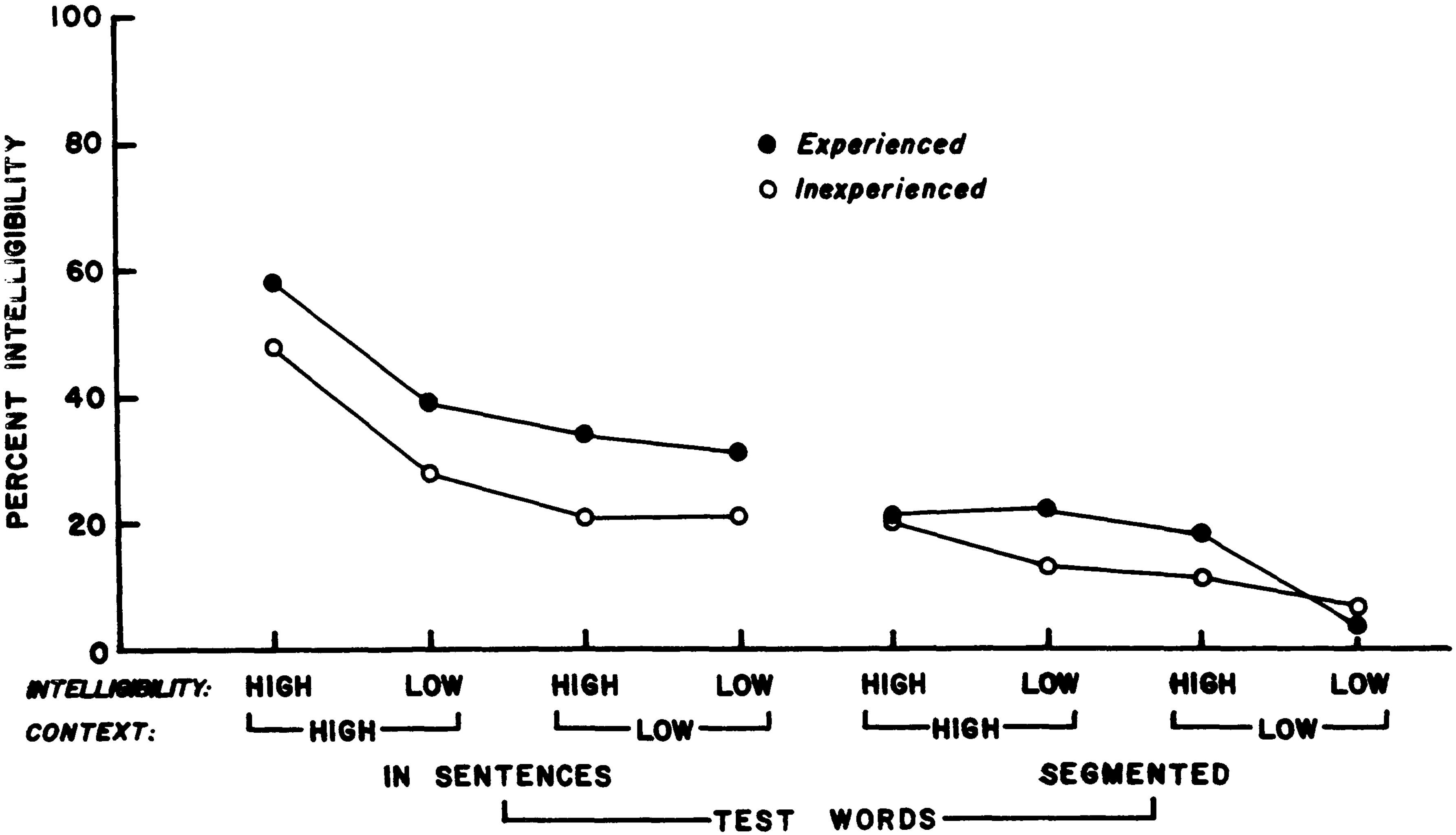

100
80
60
40
20
0
PERCENT INTELLIGIBILITY
Experienced
Inexperienced
SENTENCE
ISOLATION
SEGMENTED
SENTENCE
ISOLATION
SEGMENTED
HIGH
LOW
A PRIORI INTELLIGIBILITY

factors. The interactions between experience, context, and *a priori* intelligibility (EIC) was not statistically significant for test words in sentences but showed a borderline interactive effect for segmented test words. Scores for these three factors are summarized in Figure 9. The data are shown separately for test words in sentences and segmented test words.

For test words in sentences, the curves for experienced and inexperienced listeners are essentially parallel with the difference between listeners averaging about ten percentage points across the four combinations of *a priori* intelligibility and context. For both groups of listeners, the ranking of scores (from highest to lowest) as a function of *a priori* intelligibility and sentence context were: (1) high intelligibility, high context, (2) low intelligibility, high context, (3) high intelligibility, low context, and (4) low intelligibility, low context.

For segmented test words, the curves for experienced and inexperienced listeners were not parallel although the rankings of relative intelligibility are virtually the same for both groups; that is, for both experienced and inexperienced listeners, high context words were most intelligible and low context words, least intelligible. Also, on average, the high *a priori* intelligibility words were more intelligible than the low *a priori* intelligibility words, although for experienced listeners' high context words, the low intelligibility words were fractionally higher.

Figure 9. Mean scores obtained by experienced and inexperienced listeners for test words as a function of sentence context and *a priori* intelligibility.


PERCENT INTELLIGIBILITY
100
80
60
40
20
0
Experienced
Inexperienced
INTELLIGIBILITY: HIGH LOW HIGH LOW HIGH LOW HIGH LOW
CONTEXT: HIGH LOW HIGH LOW
IN SENTENCES
SEGMENTED
TEST WORDS

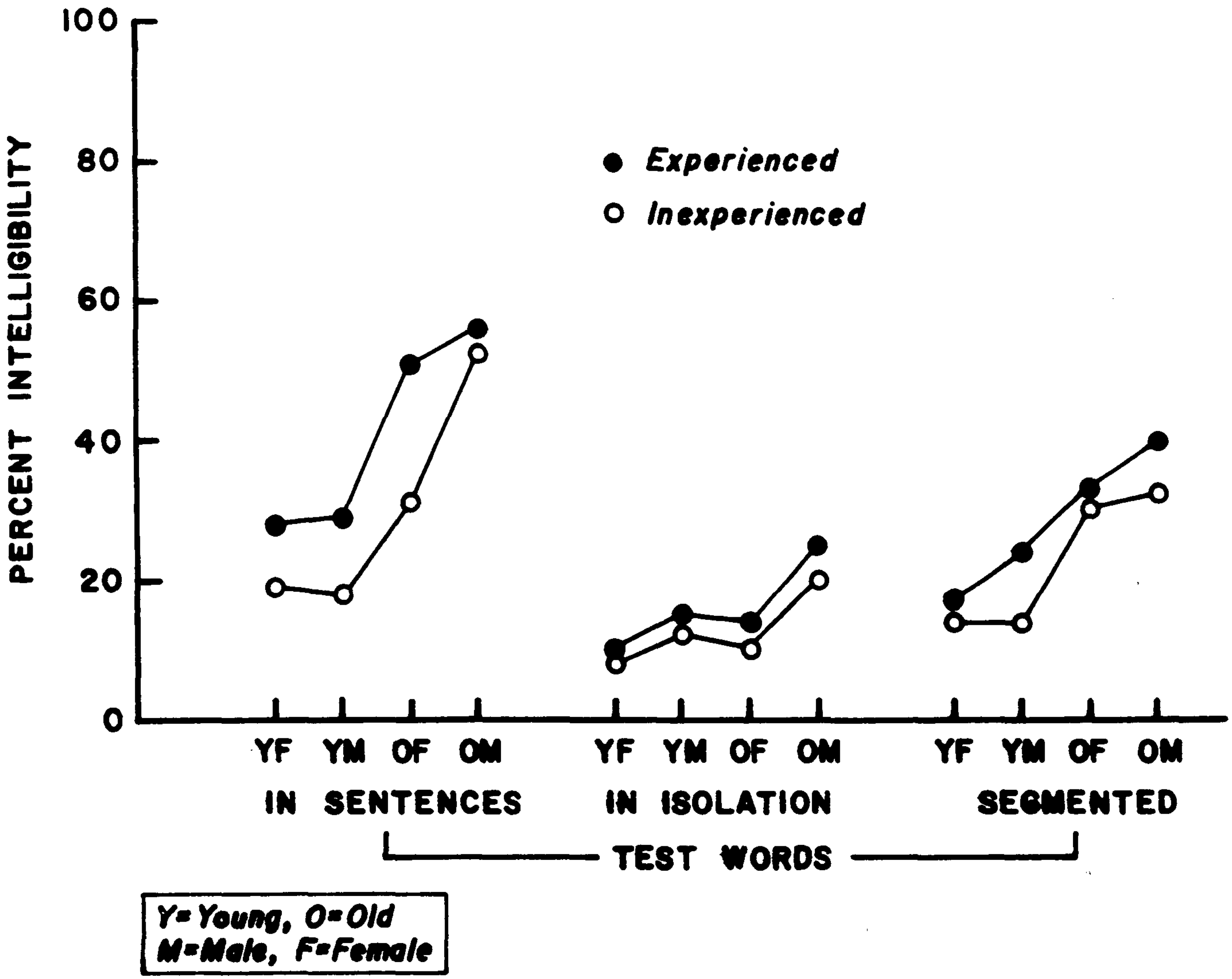
In contrast, for low context segmented test words, the experienced listeners showed a greater drop from high to low intelligibility words. This gives rise to the one case where the curves cross indicating an interaction.

Between Children Differences

Intelligibility scores were also analyzed for factors of the children's age and sex. Because age and sex differences are necessarily confounded with between subject differences, the F-ratios for these effects used the between children sum of squares as the denominator as shown in Tables 9, 10 and 11.

Figure 10 is a plot of the scores obtained by experienced and inexperienced listeners as a function of children's age and sex. The ordinate shows the percent intelligibility obtained by experienced and inexperienced listeners; the abscissa shows the groups of young and old, male and female speakers. Scores are plotted for test words in sentences, in isolation, and in segmented conditions. Overall scores for experienced and inexperienced listeners are essentially parallel indicating no interactions between listener experience and these variables. As indicated by the analysis of variance shown in Tables 9, 10, and 11, age was a significant factor for test words in sentences and in isolation, but not for segmented test words. The data of Figure 10 shows that the older children were consistently better than the younger children. This was true for all three types of stimuli. However, there were no significant

Figure 10. Mean scores obtained by experienced and inexperienced listeners for test words as a function of the children's age and sex.


PERCENT INTELLIGIBILITY
100
80
60
40
20
0
● Experienced
○ Inexperienced
YF YM OF OM
IN SENTENCES
YF YM OF OM
IN ISOLATION
YF YM OF OM
SEGMENTED
TEST WORDS
Y=Young, O=Old
M=Male, F=Female

sex differences for test words in sentences and isolation, and only a borderline significance level for segmented test words.

In order to examine individual differences between children and possible interactions with other factors, the children were ranked in terms of average intelligibility and the resulting scores plotted for experienced and inexperienced listeners for each grouping of the children. These data are shown in Figures 11, 12, and 13 for test words in sentences, in isolation, and in segmented conditions respectively. In each plot, the ordinate shows the percent intelligibility score; the abscissa shows the subject's identification code (from 1 to 20), with the children grouped according to age and sex. Within each group, the children are ranked from least to most intelligible based on the averaged intelligibility scores. This method of representing the data allows one to examine the individual differences as well as any possible interactions within relative intelligibility.

Figure 11 shows the data for test words in sentences. Within each group of children, the scores for the experienced and inexperienced listeners were roughly parallel, the higher scores being obtained by the experienced listeners. An exception to this pattern is seen with child ten for whom the scores for the inexperienced listeners were slightly higher. Two other exceptions occurred with children four and thirteen. In these cases, the scores obtained by the

Figure 11. Mean scores obtained by experienced and inexperienced listeners for test words in sentences as a function of the children's rank of intelligibility, age and sex.

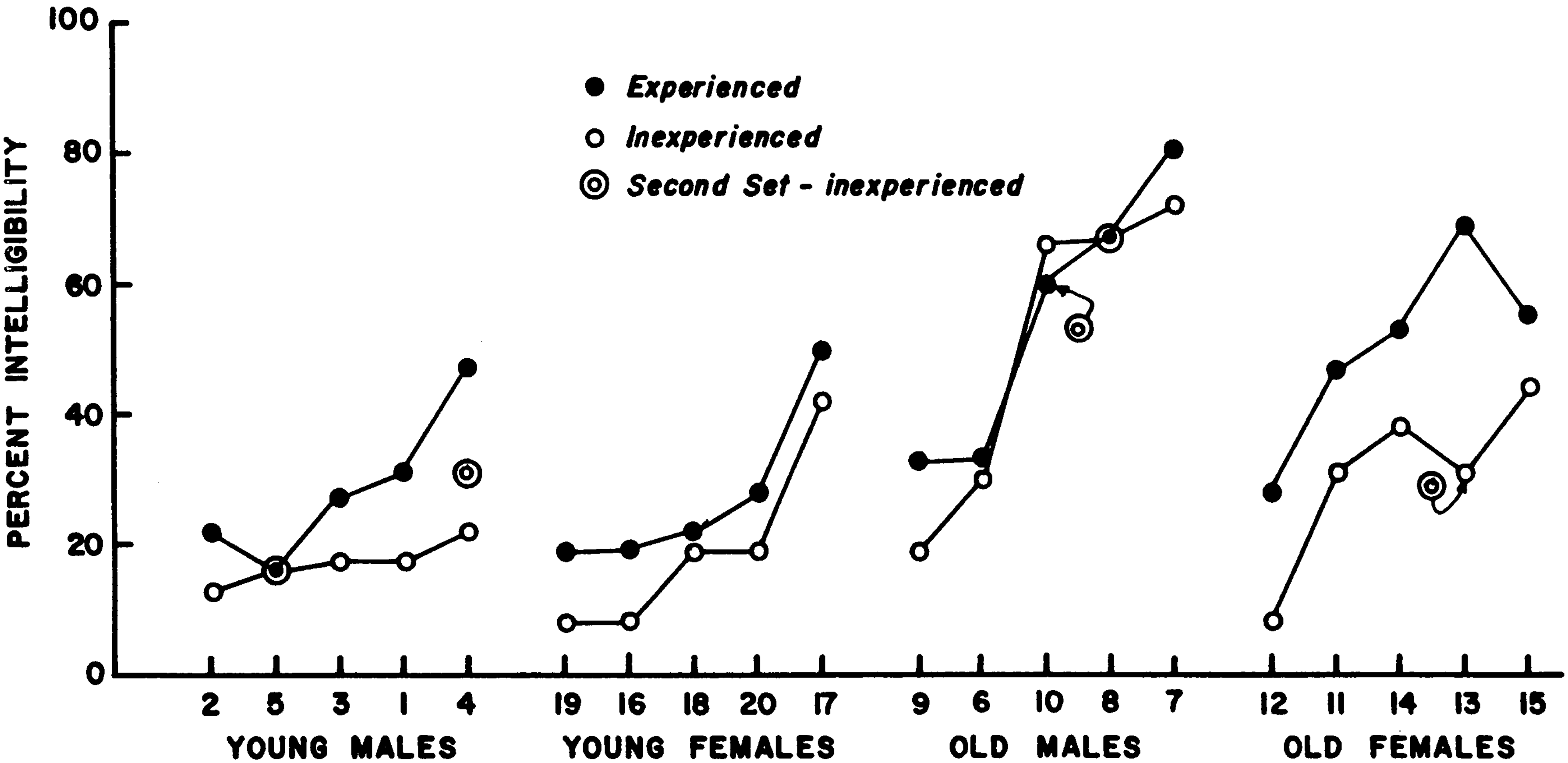

PERCENT INTELLIGIBILITY
100
80
60
40
20
0
Experienced
Inexperienced
Second Set - inexperienced
2
5
3
1
4
19
16
18
20
17
9
6
10
8
7
12
11
14
13
15
YOUNG MALES
YOUNG FEMALES
OLD MALES
OLD FEMALES

Figure 12. Mean scores obtained by experienced and inexperienced listeners for test words in isolation as a function of the children's rank of intelligibility, age and sex.

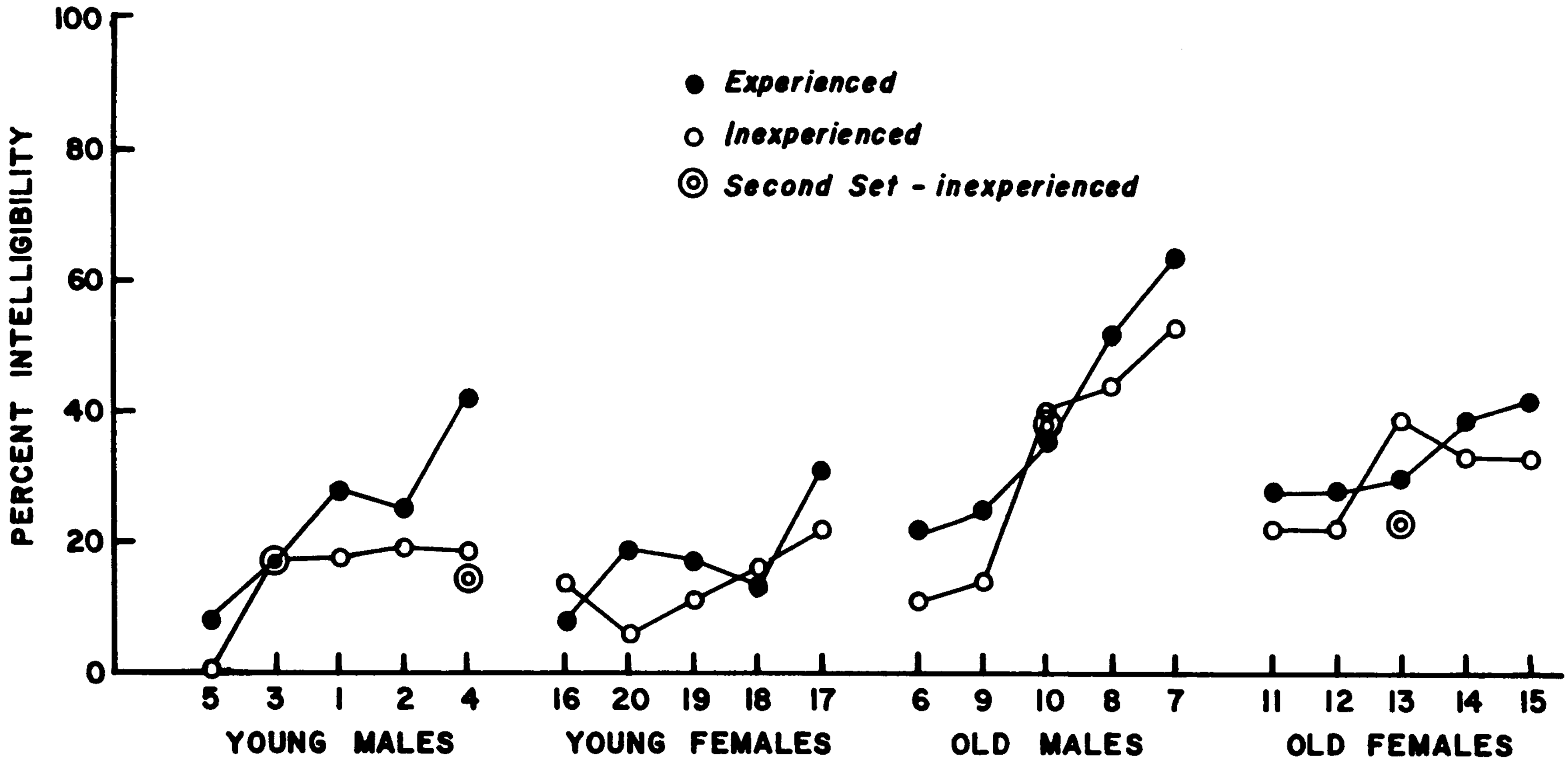
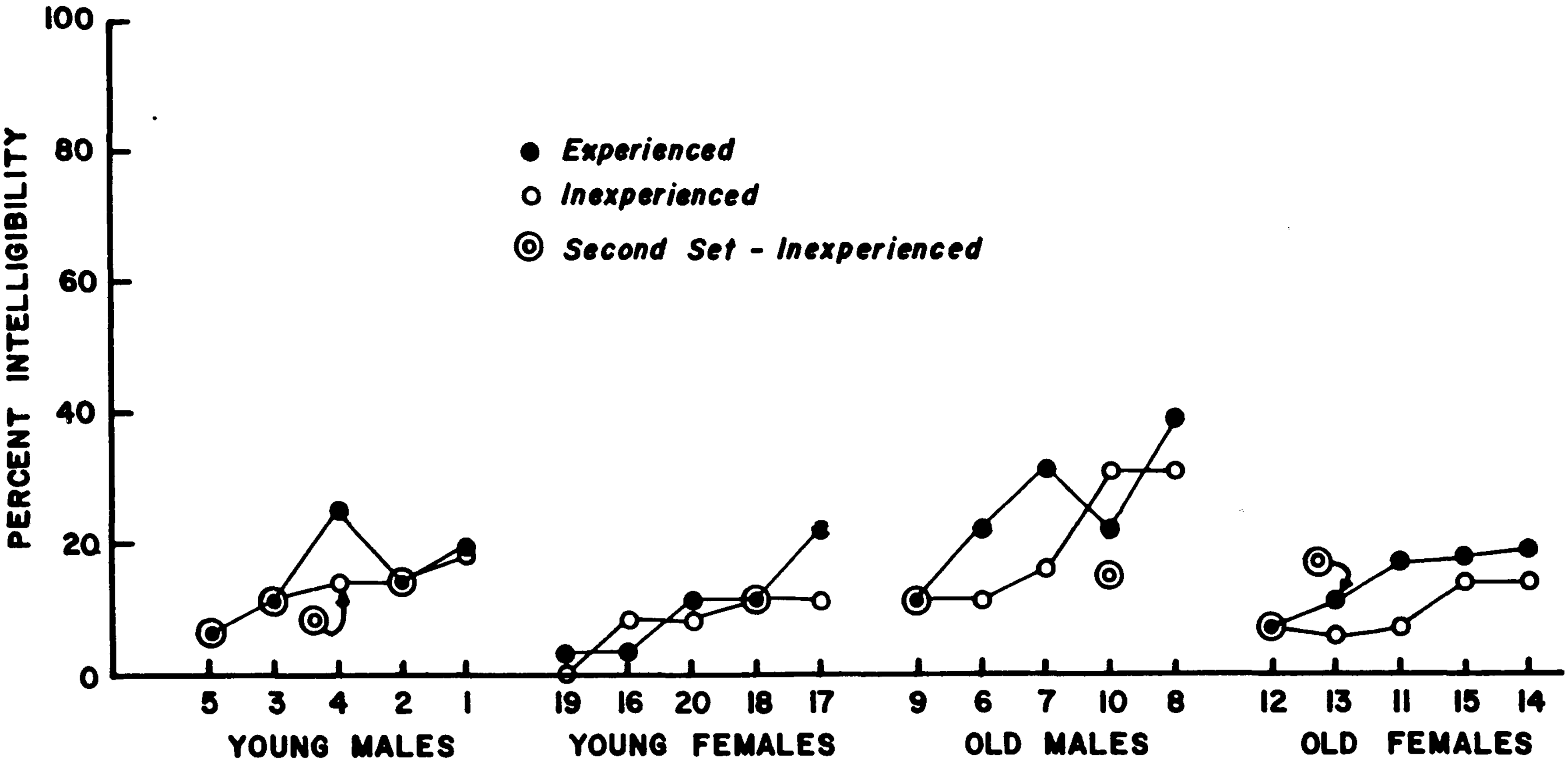

Experienced
Inexperienced
Second Set - inexperienced
PERCENT INTELLIGIBILITY
100
80
60
40
20
0
5 3 1 2 4
YOUNG MALES
16 20 19 18 17
YOUNG FEMALES
6 9 10 8 7
OLD MALES
11 12 13 14 15
OLD FEMALES

Figure 13. Mean scores obtained by experienced and inexperienced listeners for segmented test words as a function of the children's rank of intelligibility, age and sex.


PERCENT INTELLIGIBILITY
100
80
60
40
20
0
Experienced
Inexperienced
Second Set - Inexperienced
5 3 4 2 1
YOUNG MALES
19 16 20 18 17
YOUNG FEMALES
9 6 7 10 8
OLD MALES
12 13 11 15 14
OLD FEMALES

experienced listeners were considerably higher than those of the inexperienced listeners.

Figure 12 shows the data for the test words in isolation. Overall scores for experienced listeners were higher than inexperienced listeners for each group of children. In four cases (children ten, thirteen, sixteen and eighteen), scores for the inexperienced listeners were slightly better than the experienced listeners. A large difference between listeners is seen for child four, with the experienced listeners' scores higher.

The scores obtained by experienced and inexperienced listeners for segmented test words are plotted in Figure 13. Overall differences between listeners for each group of children were small. Of interest is child ten for whom the inexperienced listeners once again obtained scores higher than the experienced listeners, and child four for whom the difference between listeners continued to be large.

In summary, the overall pattern of relative intelligibility of the deaf speakers was similar for both experienced and inexperienced listeners. However, exceptions to this pattern were observed for three children. For each type of speech stimulus, the scores for child four, child ten, and child thirteen were consistently different. For child ten, scores for the inexperienced listeners were higher than those for the experienced listeners. For child four, scores for the inexperienced listeners were much poorer than the experienced group. Scores for child thirteen varied with the experimental condition. For test words in sentences, scores for inexperienced listeners were considerably poorer than experienced listeners. For isolated words, scores for inexperienced listeners were somewhat higher than experienced. For segmented words, the inexperienced listeners' scores were lower.

In each of these cases, the difference between children may have been confounded by the difference between listeners since each child was heard by three different experienced and three different inexperienced listeners. In order to check this possibility, data was obtained from a second set of inexperienced listeners. These additional scores are also plotted in Figures 11 to 13.

For child four, scores for test words in sentences obtained by the second set of inexperienced listeners were higher than those obtained by the first group, and more closely approximated the experienced listeners' scores. Scores for test words in isolation and in segmented conditions were the same or slightly poorer than those obtained for the first group of listeners.

For child ten, scores obtained from the second set of inexperienced listeners were lower for all experimental conditions than those obtained from the first set of listeners. These second set of data were the same, or slightly poorer than those obtained by the experienced listeners.

For child thirteen, scores obtained from the second set of inexperienced listeners were the same for test words in sentences, slightly poorer for test words in isolation, and slightly higher for segmented test words, than those obtained from the first set of listeners. Thus, scores for inexperienced listeners continued to be considerably poorer than experienced listeners for test words in sentences, but as expected for isolated and segmented words.

Ancillary Analysis of Listeners' Scores

An additional analysis was performed in order to check the relative variability of the intelligibility scores of the experienced and inexperienced listeners. In this analysis, the scores obtained by experienced and inexperienced listeners were first averaged for each child. Then, these averaged scores for the 20 deaf children were ranked from the most to least intelligible speaker. Separate rankings were obtained for test words in sentences, in isolation, and segmented conditions. The scores for experienced and inexperienced listeners were then plotted as a function of the children's ranks. Separate plots of sixty points (one for each listener) were made for experienced and inexperienced listeners for each experimental condition. The plots are shown in Figure 14 to 19. In each of these diagrams, the ordinate shows the percent intelligibility; the abscissa shows the rank of children. The solid line is a quadratic curve fitted by the method of least squares to the measured intelligibility scores. The deviation between the plotted points and the fitted line is made up of these components:

Figure 14. Scores obtained by experienced listeners for test words in sentences plotted as a function of the children's rank of intelligibility.

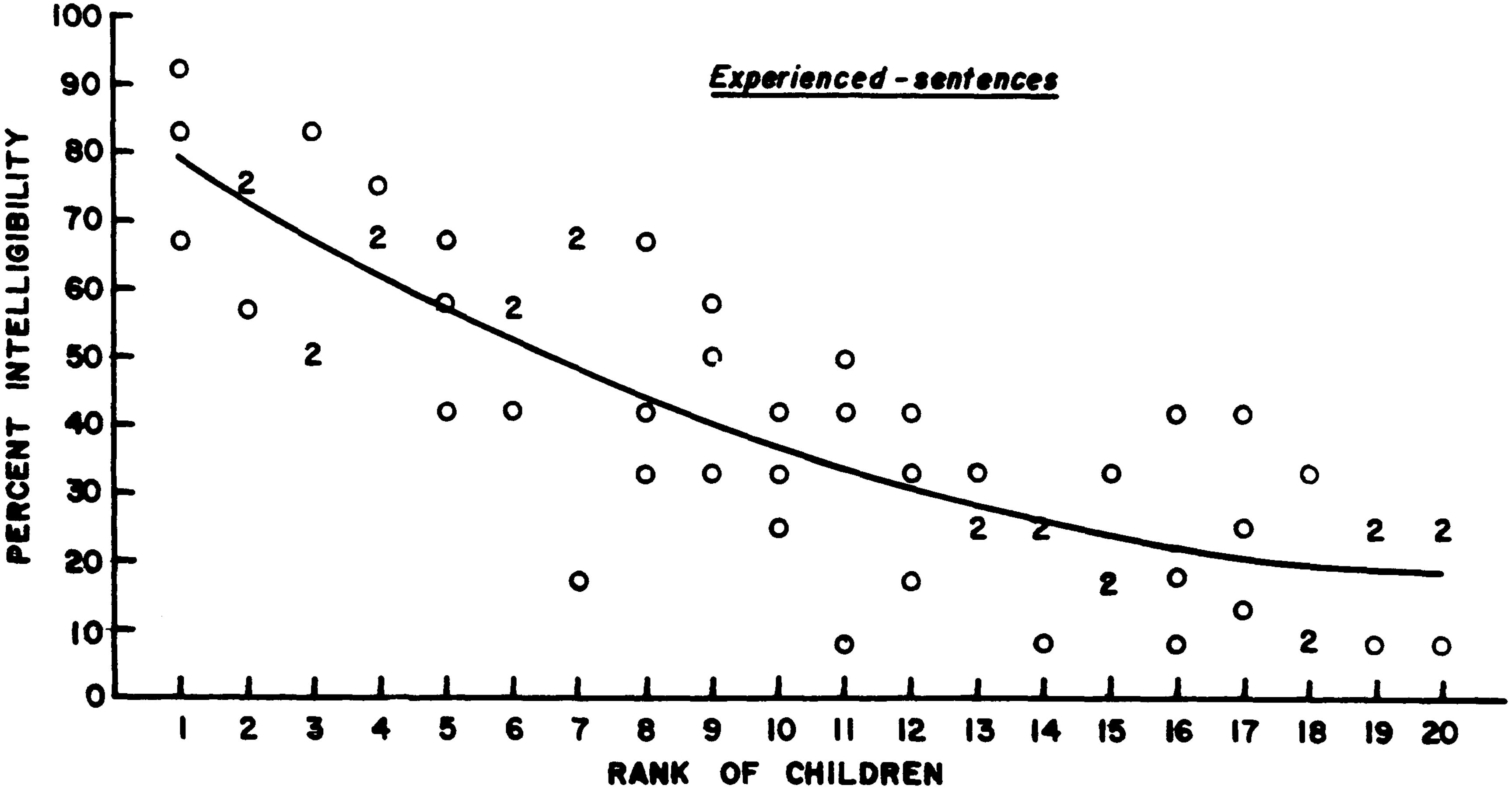
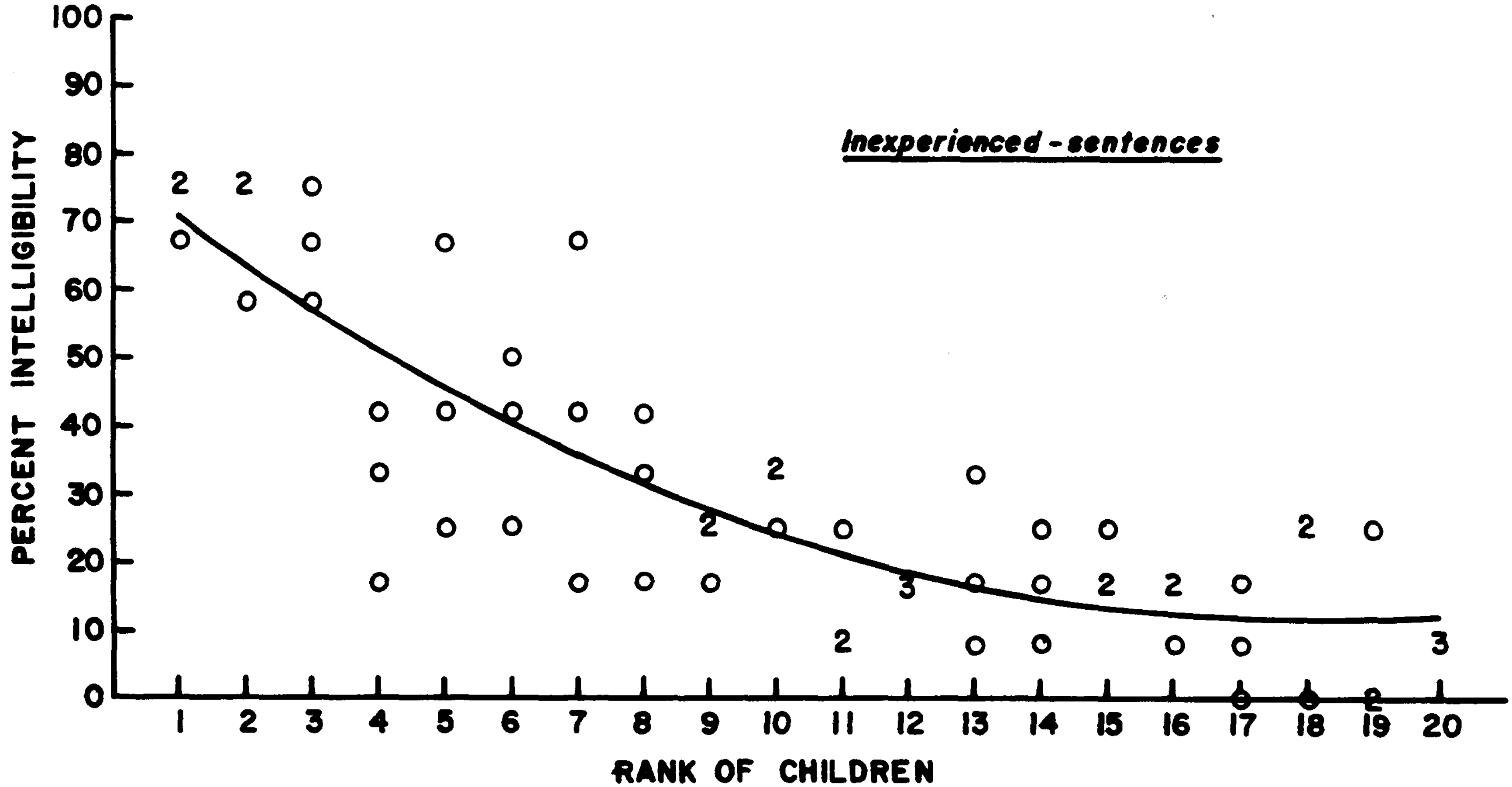

Experienced - sentences
PERCENT INTELLIGIBILITY
0
10
20
30
40
50
60
70
80
90
100
RANK OF CHILDREN
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figure 15. Scores obtained by inexperienced listeners for test words in sentences plotted as a function of the children's rank of intelligibility.



PERCENT INTELLIGIBILITY
RANK OF CHILDREN
Inexperienced - sentences

Figure 16. Scores obtained by experienced listeners for test words in isolation plotted as a function of the children's rank of intelligibility.

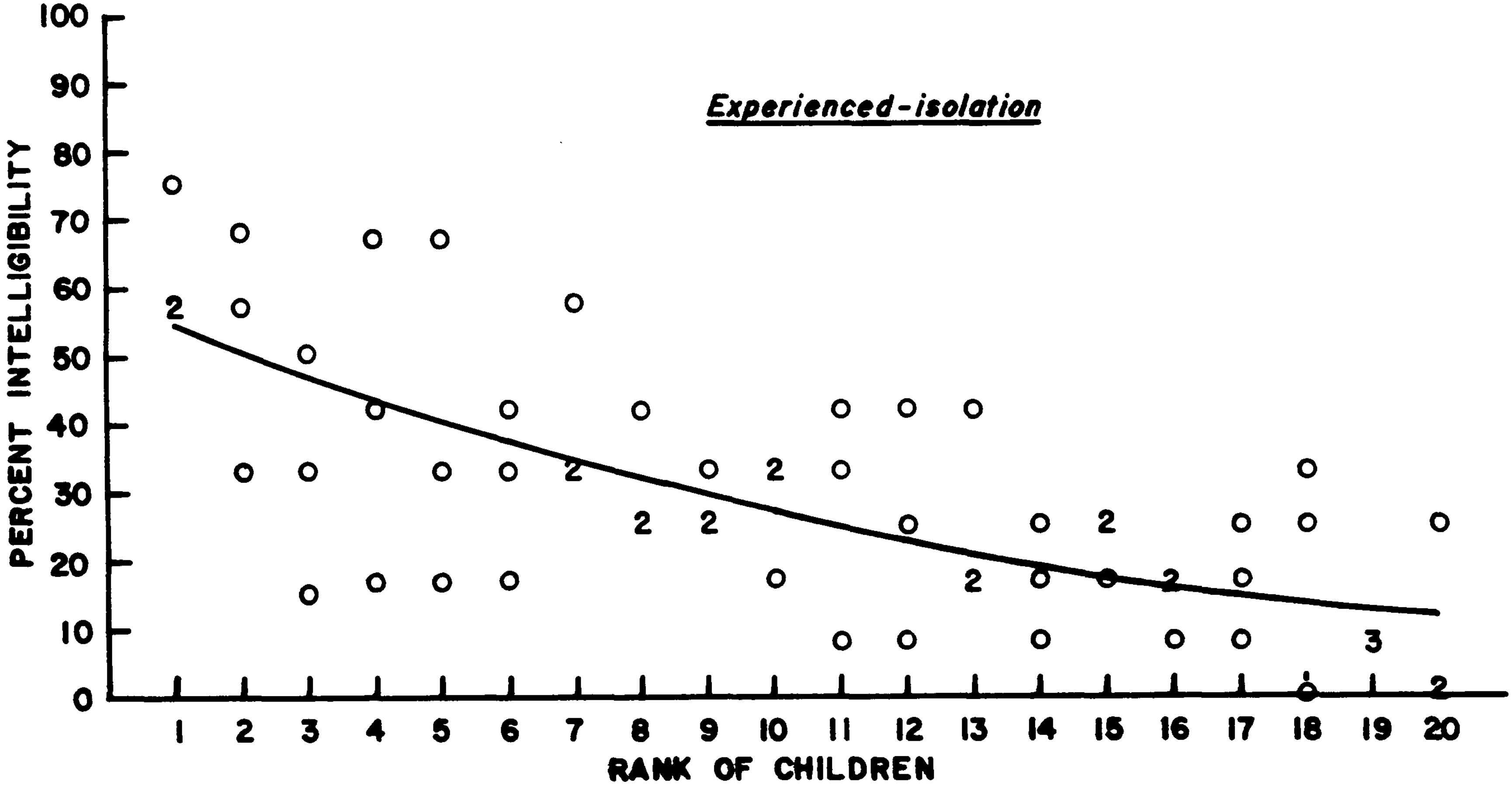

Experienced-isolation
PERCENT INTELLIGIBILITY
100
90
80
70
60
50
40
30
20
10
0
RANK OF CHILDREN
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figure 17. Scores obtained by inexperienced listeners for test words in isolation plotted as a function of the children's rank of intelligibility.

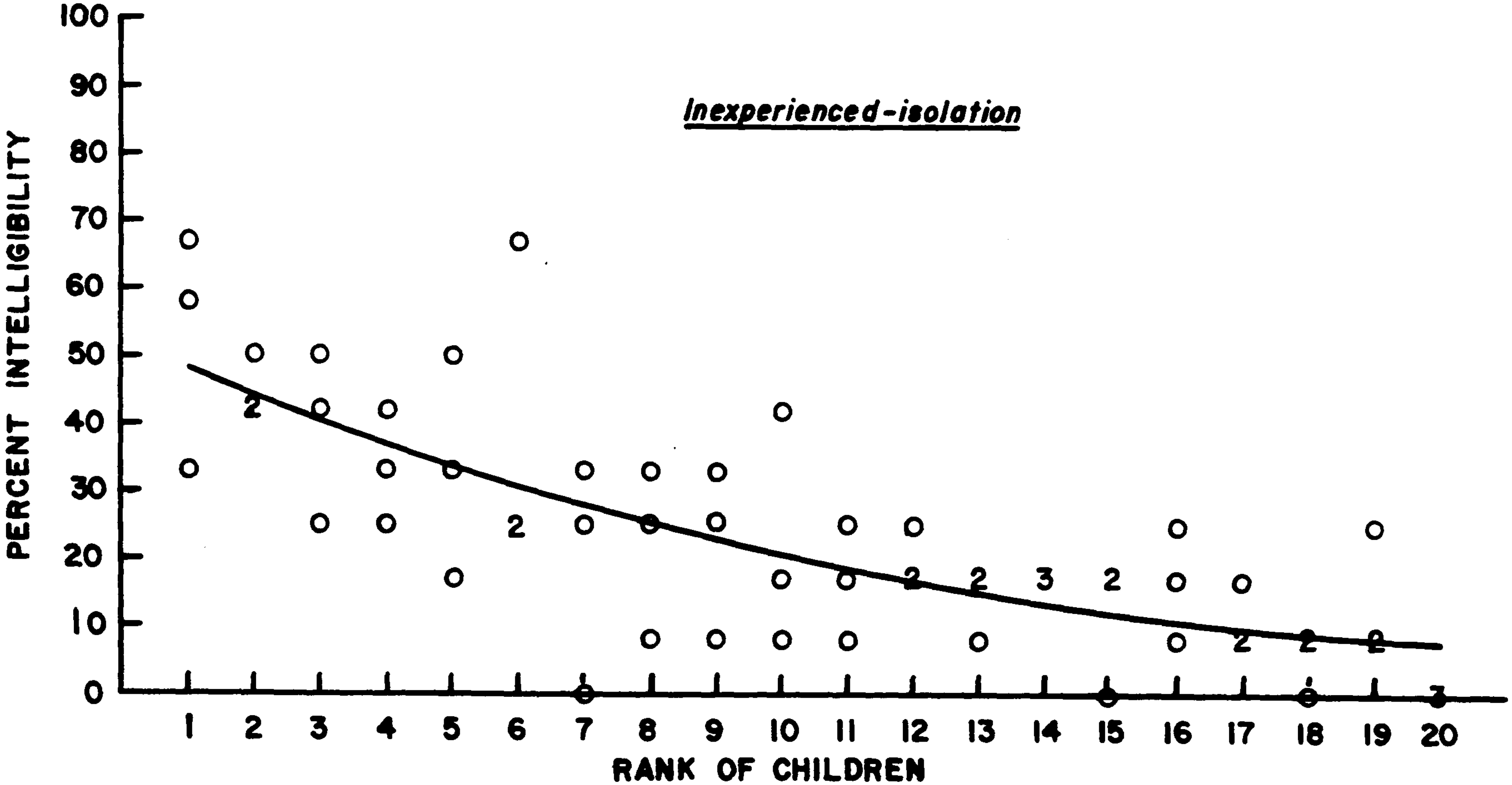

Inexperienced-isolation
PERCENT INTELLIGIBILITY
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
RANK OF CHILDREN

Figure 18. Scores obtained by experienced listeners for segmented test words plotted as a function of the children's rank of intelligibility.

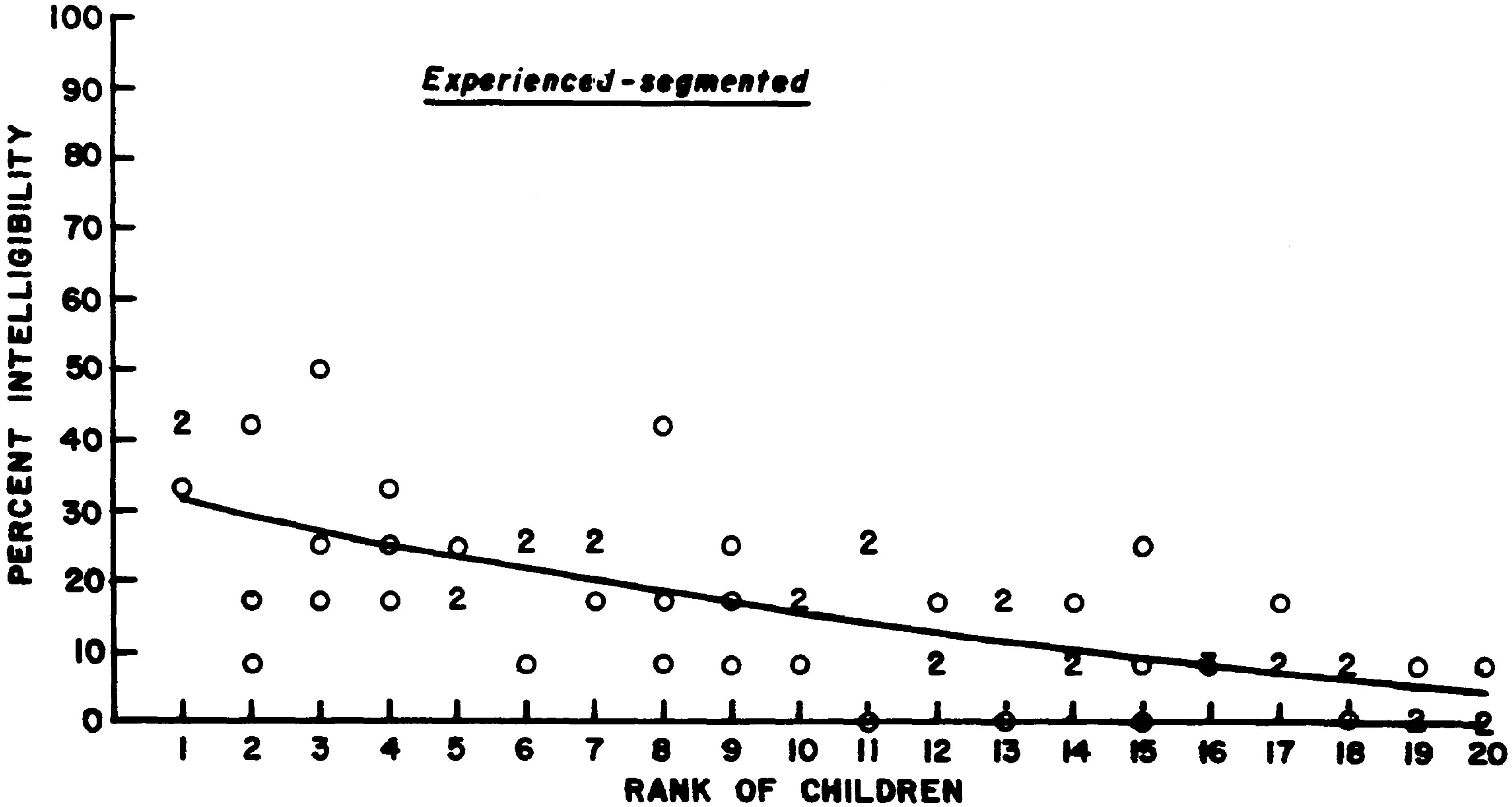
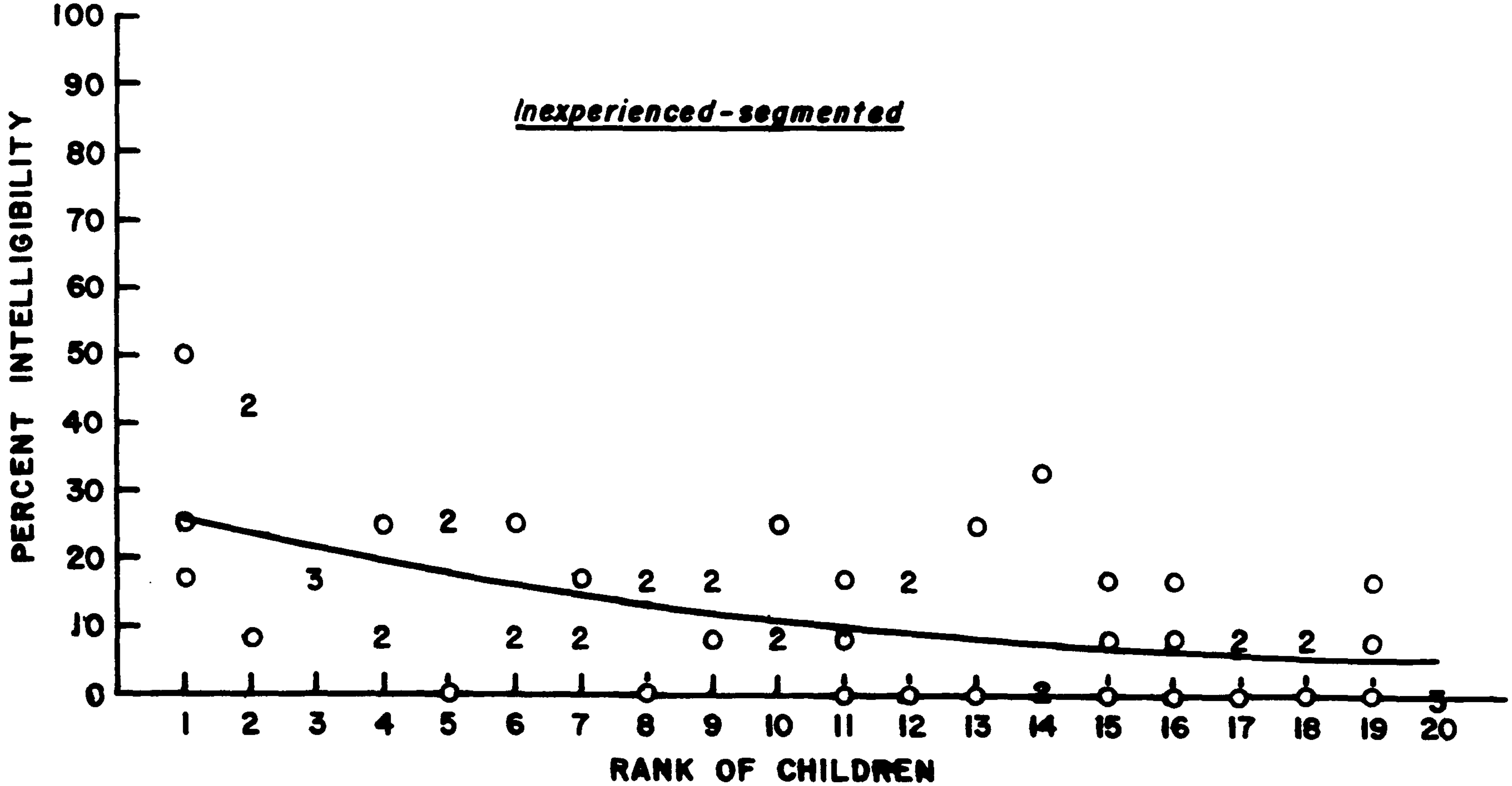

Experienced-segmented
PERCENT INTELLIGIBILITY
RANK OF CHILDREN

Figure 19. Scores obtained by inexperienced listeners for segmented test words plotted as a function of the children's rank of intelligibility.


Inexperienced-segmented
PERCENT INTELLIGIBILITY
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
RANK OF CHILDREN

(1) between-listener differences (because each point represents a different listener), (2) test-retest differences, and (3) any deviations between the shape of the true curve relating percent intelligibility to children's rank, and the shape of the fitted curve.

Of the above components, test-retest differences are typically smaller than between listener differences. Also, deviations due to point three are expected to be small since the method of ranking necessarily produced a smooth monotonic function, which is well fitted by the quadratic curve. Further, any large systematic differences from the fitted curve will be evident from these diagrams. Thus, the deviations from the fitted curve are believed to be due primarily to between-listener differences.

Figures 14 and 15 are the plots for the test words in sentences, for experienced and inexperienced listeners respectively. Scores for the experienced listeners generally cluster close to the fitted curve (standard deviation = 0.28 in arc sine units). Similarly, scores for inexperienced listeners, while somewhat lower than those for the experienced listeners, also lie in close proximity to the fitted curve. (standard deviation = 0.31 in arc-sine units). The curve covers the range from roughly 75% to 15%.

The plots for test words in isolation are shown in Figure 16 for the experienced, and Figure 17 for the inexperienced listeners. For both groups of listeners, the fitted curves are shallower than their respective curves for sentences. In this case, the fitted curves cover a range from roughly 51% to 10%. The deviation of the plotted points from the fitted line was essentially the same as in the previous case; the standard deviations being 0.32 arc – sine units for both experienced and inexperienced listeners.

The plotted points for segmented test words are shown in Figures 18 and 19, for both experienced and inexperienced listeners respectively. These curves show a further reduction in range, from a maximum of roughly 29%, to a minimum of 5%. The standard deviation of the plotted points from the fitted line is essentially the same for the experienced listeners (standard deviation = 0.28 arc-sine units). There is, however, a slight increase in variability for the inexperienced listeners, the standard deviation in this case being 0.36 arc-sine units. Inspection of the data indicates that this increase in standard deviation is due to many more listeners obtaining scores of zero. These points differed from the fitted line by a larger than average amount.

In summary, there was no consistent difference between experienced and inexperienced listeners in terms of relative variability, except in the special case when intelligibility was low and more inexperienced listeners obtained scores of zero.

Effect of the Position of the Test Word in the Sentence and the Number of Syllables

An additional analysis of variance was performed to investigate the effect of the position of the test words in the sentence, the number of syllables in the sentence, and whether there were any interactions between listener experience and these two factors. This analysis of variance appears in Table 14. In order to perform this analysis, it was necessary to collapse the data over other non-interactive factors in order to obtain a large enough number of samples for the intelligibility scores. Inspection of the data showed no evidence of second or third order interaction between context, _a priori_ intelligibility, and the factors of number of syllables or position of the test word in the sentence. Similarly, there was no strong evidence of an interaction between children and the latter two factors. For these reasons, it was decided to collapse the data over _a priori_ intelligibility, context, and children in the subsequent analysis.

The factors considered in this final analysis were: listener experience, number of syllables, position of the test words, and type of stimulus. Table 14 summarizes the analysis of variance. For technical reasons, the analysis of variance was done without the arc-sine transformation. However, the F ratios were not expected to be different. As before, because of the large number of F tests, only those factors with a significance level of .01 or smaller were considered significant.

Of particular interest was the effect of the number of syllables in the sentence, the position of the test word,

TABLE 14

ANALYSIS OF VARIANCE FOR NUMBER OF SYLLABLES AND WORD POSITION*

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
Number of Syllables (N). . . .	1.17	2	0.58	3.46	.031
Position of Test Word (P). . .	3.14	2	1.57	9.26	.000
Type of Stimulus (T)	31.45	1	31.45	185.54	.000
Listener Experience (E). . . .	3.82	1	3.82	22.57	.000
N x P.	6.21	4	1.55	9.16	.000
N x T.	.73	2	.36	2.16	.115
N x E.	.10	2	.05	.30	.740
P x T.	.43	2	.21	1.29	.275
P x E.	.19	2	.09	.55	.572
T x E.	.97	1	.97	5.75	.017
N x P x T.	3.28	4	.82	4.84	.001
N x P x E.	1.33	4	.33	1.96	.097
N x T x E.	.60	2	.30	.78	.168
P x T x E.	.04	2	.02	.73	.872
N x P x T x E.	.53	4	.13	.79	.529
Explained.	54.07	35	1.54	9.11	.000
Residual	482.20	2844	.170	----	----
Total.	536.27	2879	.186	----	----

* For technical reasons, this analysis of variance was done without the arc sine transformation.

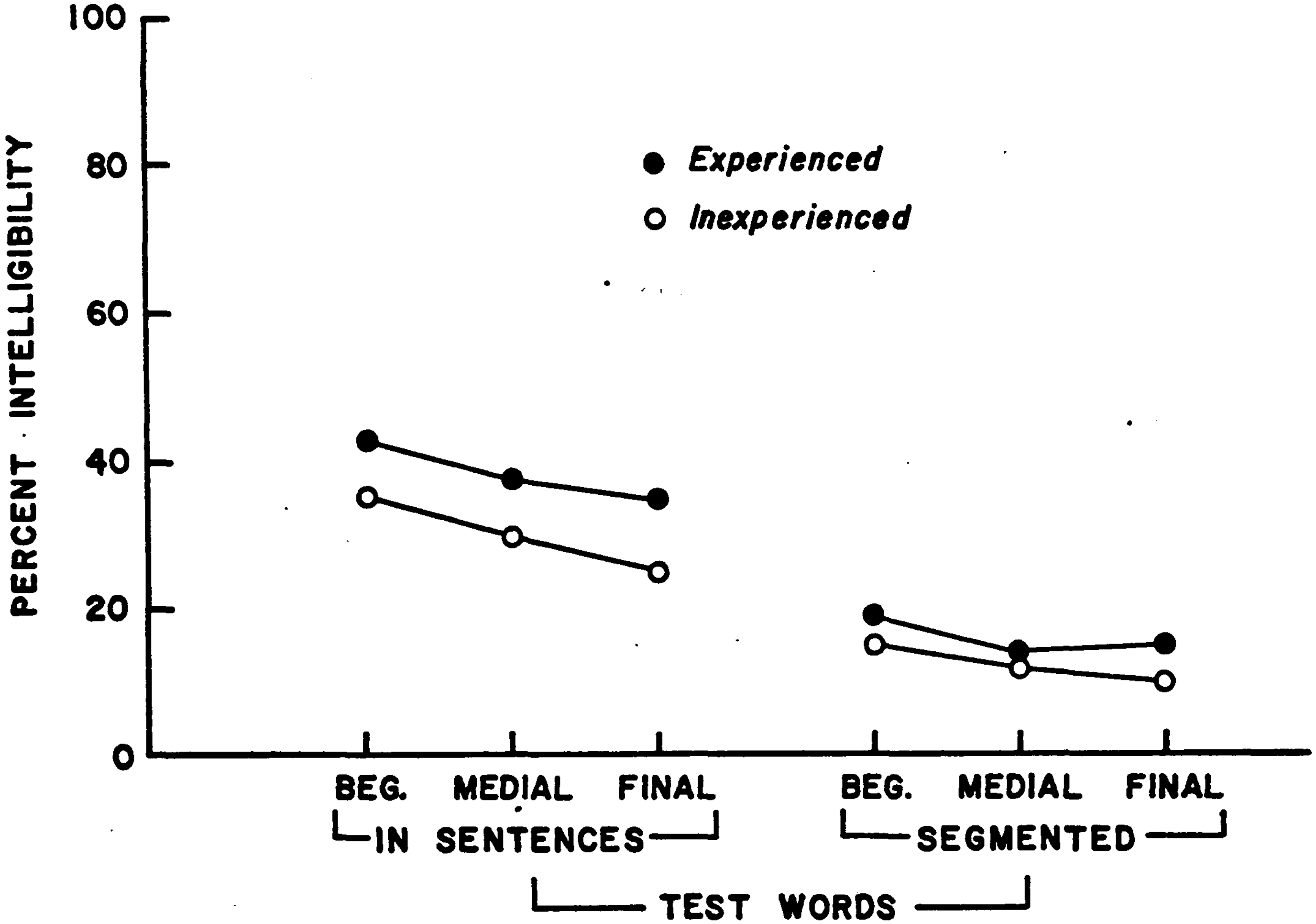
and any interactions involving these factors. Table 14 shows that the main effect for the position of the test word was highly significant. No statistical effect was found for the number of syllables in the sentence. However, there was a significant interaction between the number of syllables and the position of the test word in the sentence (NxP). There was also some evidence of an interaction between the type of stimulus and listener experience (TxE). Of the three-way and higher order interactions, only one was found to be significant. This was the three way interaction between number of syllables, position of the test word, and type of stimulus (NxPxT).

Scores for the position of the test word in the sentence are shown in Figure 20 for both experienced and inexperienced listeners. The percent intelligibility is plotted on the ordinate; the position of the test word (that is, beginning, middle or end) is plotted on the abscissa. The left hand side of the diagram shows the data for sentences; the right hand side, the data for segmented words.

On average, the experienced listeners obtained higher scores than the inexperienced listeners for all three positions of test words whether in sentences or segmented conditions.

For test words in sentences, the pattern of relative intelligibility was similar for both groups of listeners. Scores were highest for test words in the beginning of sentences, followed by those in the middle, which were in

Figure 20. Mean scores obtained by experienced and inexperienced listeners for test words as a function of the position in the sentence.


100
80
60
40
20
0
PERCENT · INTELLIGIBILITY
Experienced
Inexperienced
BEG.
MEDIAL
FINAL
IN SENTENCES
BEG.
MEDIAL
FINAL
SEGMENTED
TEST WORDS

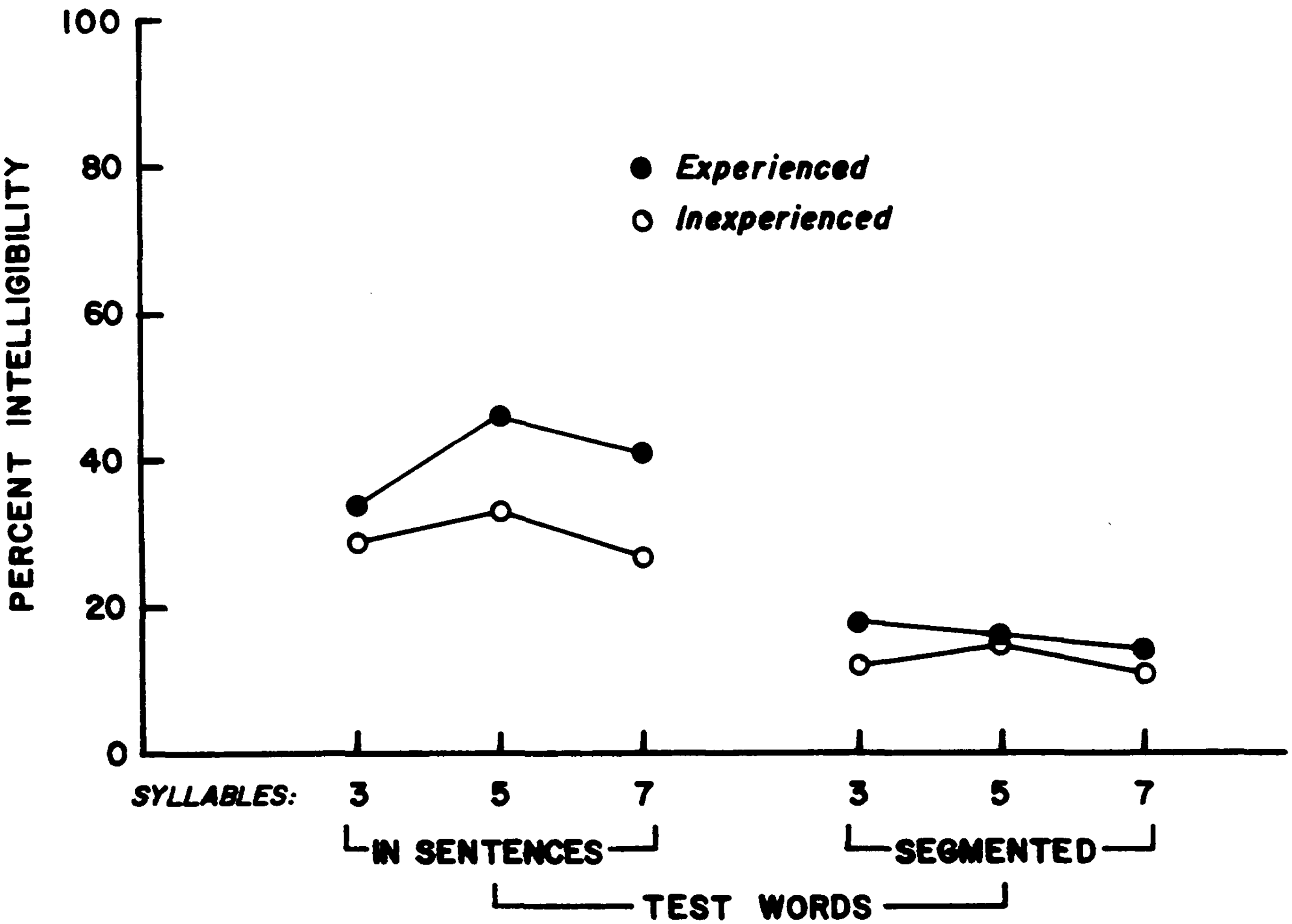
turn higher than those at the end of sentences. The difference between experienced and inexperienced listeners was roughly ten percentage points across the three word positions.

In contrast, scores for segmented test words were only slightly higher for the experienced listeners. The difference between listeners was close for all word positions; in this case, approximately five percentage points. Scores for test words segmented from the beginning, middle, or end of the sentences showed essentially similar scores.

These plots indicate there is no significant interaction between listener experience and position of the test word. However, there is some evidence of an interaction between listener experience and type of stimulus, the effect of experience being smaller for segmented words as already described.

Figure 21 shows the scores for experienced and inexperienced listeners as a function of the number of syllables. For test words in sentences, scores for experienced listeners were higher than those obtained by inexperienced listeners for three, five, or seven syllable sentences. For experienced listeners, scores for sentences were higher for five syllable sentences, followed by seven syllable sentences, which were higher in turn than three syllable sentences. For inexperienced listeners, scores for sentences were approximately the same, with the five syllable sentences only slightly higher.

Figure 21. Mean scores obtained be experienced and inexperienced listeners for test words as a function of the number of syllables in the sentence.


100
80
60
40
20
0
PERCENT INTELLIGIBILITY
Experienced
Inexperienced
SYLLABLES:
3
5
7
3
5
7
IN SENTENCES
SEGMENTED
TEST WORDS

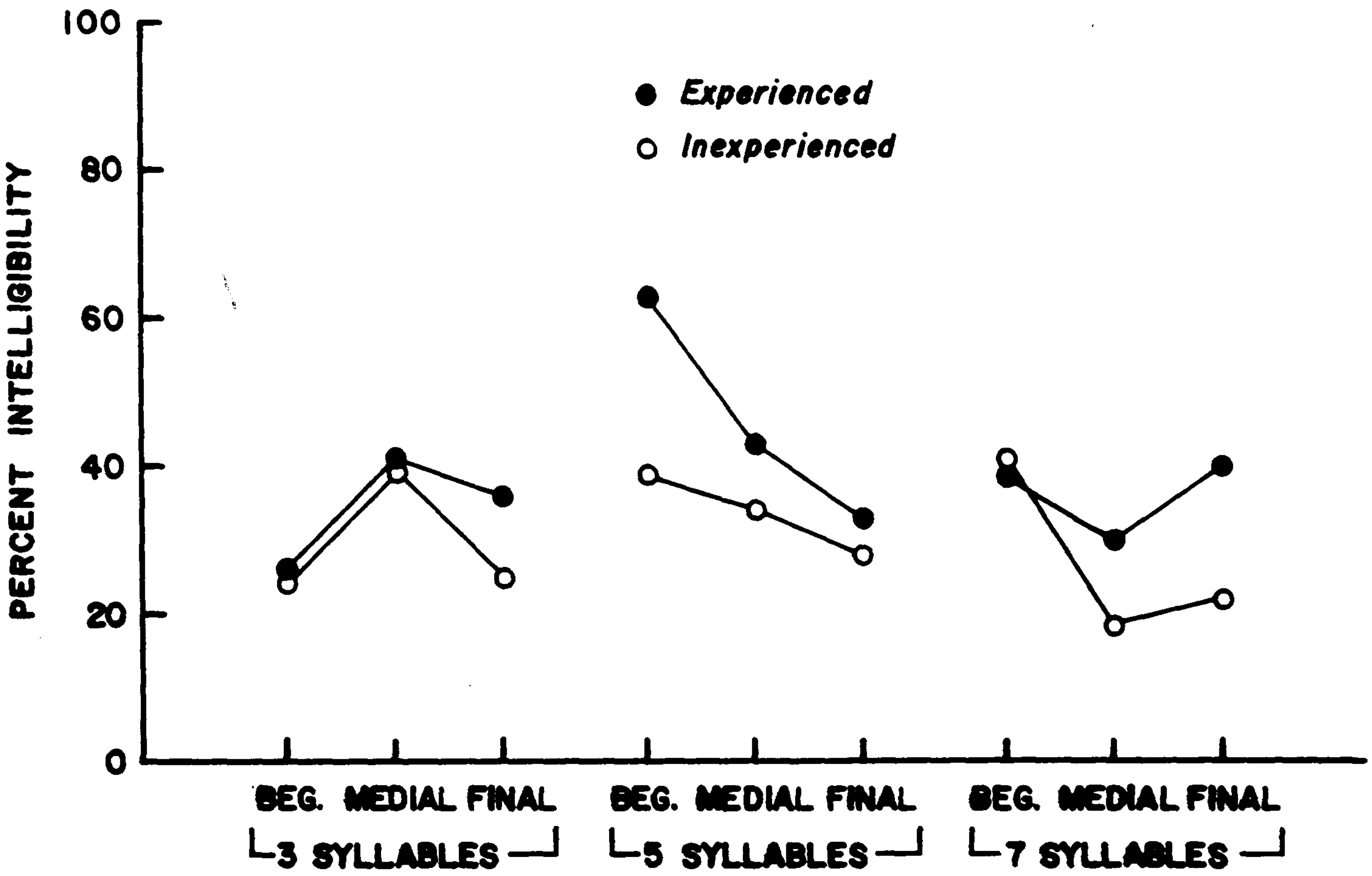
Differences between the two groups of listeners were essentially the same for five and seven syllable sentences, being on the order of fourteen percentage points. However, the difference for three syllable sentences was smaller, only five percentage points.

As already discussed, scores between experienced and inexperienced listeners were much smaller for the segmented test words. In this case however, the largest difference occurred with the three syllable sentences. However, the interaction between listener experience, number of syllables, and stimulus type was not statistically significant.

Table 12 shows that the interaction between the number of syllables in the sentence and the position of the test word was statistically significant, as was the interaction between these two factors and the type of stimulus. Both interactions can be seen in Figures 22 and 23.

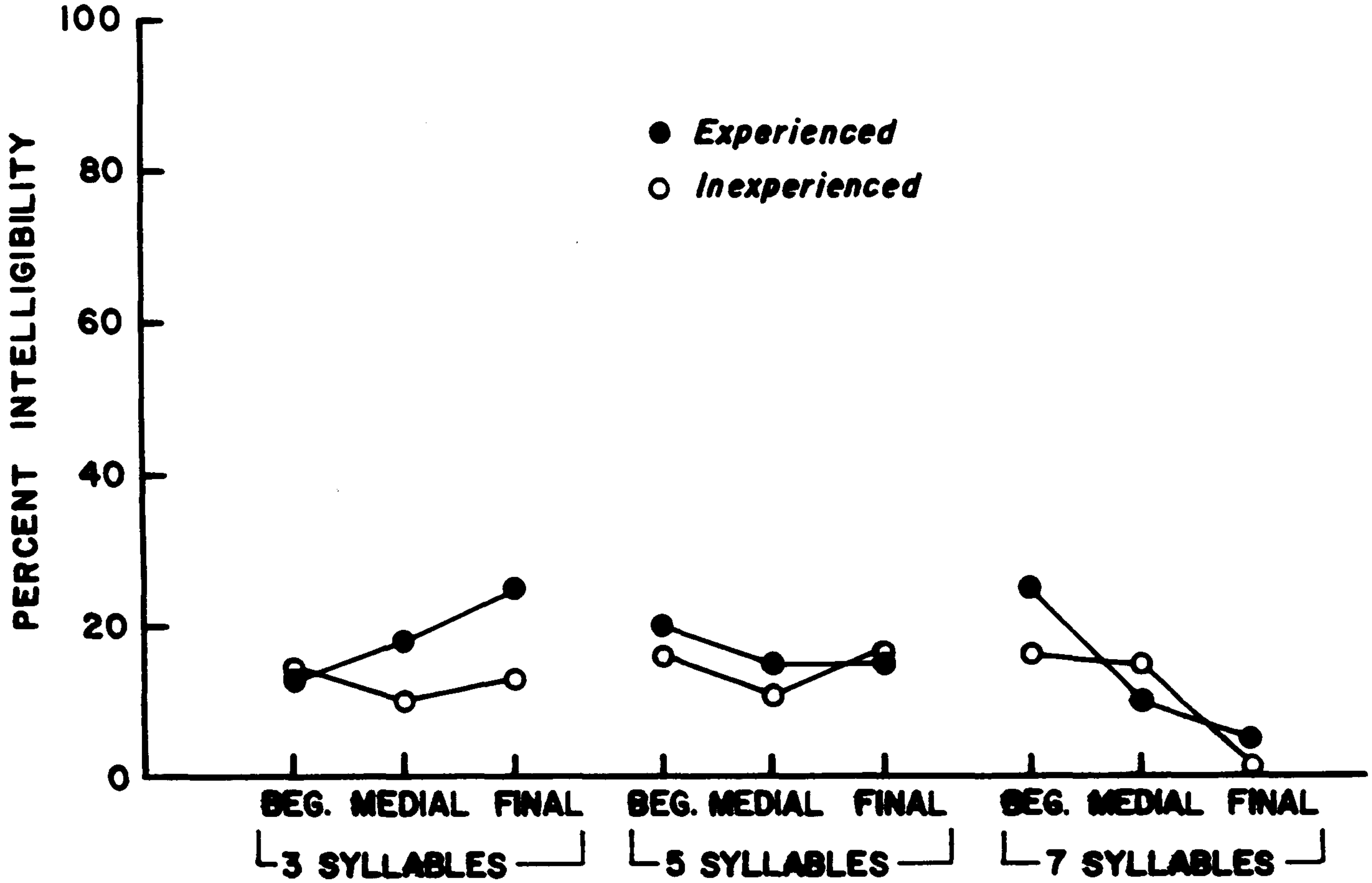
Figure 22 shows the percent intelligibility obtained by experienced and inexperienced listeners for test words in sentences as a function of number of syllables and word position. For three syllable sentences, test words at the beginning of the sentence were less intelligible than those at the beginning of five and seven syllable sentences. It should be noted that the test words in three syllable sentences were always in the word initial position, while those in the five and seven syllable sentences occurred close (within two syllables) to the beginning of the sentence but not necessarily in the word initial position.

Figure 22. Mean scores obtained by experienced and inexperienced listeners for test words in sentences as a function of word position and the number of syllables.



100
80
60
40
20
0
PERCENT INTELLIGIBILITY
Experienced
Inexperienced
BEG. MEDIAL FINAL
3 SYLLABLES
BEG. MEDIAL FINAL
5 SYLLABLES
BEG. MEDIAL FINAL
7 SYLLABLES

Figure 23. Mean scores obtained by experienced and inexperienced listeners for segmented test words as a function of word position and the number of syllables.


PERCENT INTELLIGIBILITY
100
80
60
40
20
0
Experienced
Inexperienced
BEG. MEDIAL FINAL
3 SYLLABLES
BEG. MEDIAL FINAL
5 SYLLABLES
BEG. MEDIAL FINAL
7 SYLLABLES

Differences between experienced and inexperienced listeners were greatest for test words near the beginning of five syllable sentences, and for test words near the middle and end of seven syllable sentences. However, the analysis of variance showed no statistically significant interaction between listener experience and these other factors.

Figure 23 shows intelligibility scores for experienced and inexperienced listeners for segmented test words as a function of the number of syllables and the position of the test word. As already noted, intelligibility scores for segmented test words were considerably poorer than those for test words in sentences, and also that scores for experienced and inexperienced listeners were close. Other than the observation that experienced listeners had slightly higher scores, there was no obvious pattern between listeners as a function of the number of syllables or word position.

For segmented test words, there was some evidence of a change in average scores for all listeners as a function of the number of syllables in the sentence. For the three syllable sentences, average scores increased from the beginning to the end positions. For seven syllable sentences, there was evidence of a systematic decrease from the beginning to the end of the sentence. For five syllable sentences, there was no evidence of change.

Summary of Results for Deaf Children

The data revealed several findings with respect to the differences between experienced and inexperienced listeners. These were:

1. Scores for experienced listeners were consistently higher than those for inexperienced listeners for all experimental conditions.

2. The magnitude of the difference between experienced and inexperienced listeners was largest for test words in sentences, about two thirds as large for test words in isolation, and roughly half again as large for words produced in context but heard in isolation (i.e. segmented test words).

3. There were no statistically significant interactions between listener experience and other factors considered in this study, specifically degree of sentence context, the _a priori_ intelligibility of the test word, the number of syllables in the sentence, and the position of the test word in the sentence. Similarly, there were no interactions between listener experience and children (that is, male or female, younger or older, more or less intelligible).

4. For both groups of listeners, scores for test words in sentences were higher than those for test words in isolation, which in turn were higher than segmented test words. The magnitude of the difference between test words in sentences and segmented test words was substantially greater than the differences observed between experienced and inexperienced listeners.

5. For both experienced and inexperienced lis-

teners, test words in high sentence context were more intelligible than test words in low sentence context.

6. Scores for both experienced and inexperienced listeners were greater for words with high a priori intelligibility than those with low a priori intelligibility.

7. The intelligibility scores of experienced and inexperienced listeners did not change as the overall length of the sentence increased (that is, from three to five, to seven syllable sentences).

8. There was a significant difference for words in the beginning, middle, or end of the sentence. Scores for test words in the word initial position of three syllable sentences were consistently lower than those test words occurring near the beginning of five or seven syllable sentences. There was no major differences as a function of word position for five or seven syllable sentences.

9. There was some evidence of a three way interaction between word position, number of syllables, and type of stimulus.

10. Older children were slightly more intelligible than younger children, but there was no significant difference between male and female subjects. Finally, there were no interactions between these two factors, or the relative intelligibility of the children, with any of the other variables considered.

Results For Hearing Children

Data was also collected on a group of normal speaking, hearing children. The purpose of testing this control group was to check on the relative intelligibility of each type of speech stimulus, and specifically the segmented test words, and to also compare overall findings to those obtained by the deaf children.

Eight hearing children were tested. The listeners who participated in this part of the study may be considered inexperienced in the sense that they had not previously participated in a study to evaluate either deaf or normal speech. A listing of the children's intelligibility scores obtained by these listeners for each of the experimental conditions is shown in Appendix F.

An analysis of variance was not performed for this data since a large number of the scores obtained as a function of the different factors, were nearly 100%. Data is therefore reported in percentages, obtained without the arc sine transformation.

Of primary interest was whether test words in sentences, in isolation, and in segmented conditions would have similar relative intelligibility for the hearing children as for the deaf. Intelligibility scores obtained for each type of speech stimulus were: 98% for test words in sentences, 92% for test words in isolation, and 71% for segmented test words. Scores for test words in sentences and in isolation were similar, while those for segmented test words were less intelligible. The ranking of intelligibility from test words in sentences, to isolation, to segmented conditions was the same as the ranking obtained for the deaf children. However, the difference in test scores between the types of stimuli were not as large as those observed for deaf subjects.

Data were also analyzed for test words as a function of the degree of sentence context. For test words in sentences, the scores for high context sentences were 99%, for low context sentences 97%. As previously noted, segmented test words were far less intelligible than those in sentence context. The scores for test words segmented from high context sentences was 81%, for test words segmented from low context sentences, 61%. Thus, the effect of context was slight for test words produced and heard in sentences. However, if the same words were segmented from the sentence and presented to the listener in isolation (i.e. segmented), the effect of context was substantial. This result differs from that obtained for deaf children. For the latter group, experienced and inexperienced listeners' scores showed a substantial difference between test words in high and low context sentences, but a much smaller context effect for segmented test words.

There appears to be no consistent differences between test words with high or low a priori intelligibility or between intelligibility and the type of stimulus for hearing children. The scores for test words in sentences were 99% for high intelligibility and 97% for low intelligibility words. Scores for high and low _a priori_ intelligibility words in isolation were 93% and 90% respectively. Scores for segmented test words with high _a priori_ intelligibility were 69% and 72% for those with low _a priori_ intelligibility.

Since data for deaf children showed an interaction between _a priori_ intelligibility and context, it was of interest to check if a similar interaction existed for hearing children. As can be seen in Table 15 scores for sentences were uniformly close to 100%, hence it was not possible to make meaningful comparisons for this type of speech material. However, scores for segmented words were significantly lower than 100% for hearing children. There appears to be an interaction between _a priori_ intelligibility and context that was similar in form to that obtained for deaf children, i.e. scores for test words segmented from low context sentences were consistently lower than those segmented from high context sentences. Specifically, for high context materials the scores for low intelligibility segmented words were marginally lower than the high intelligibility words. However, for test words segmented from low context sentences, the scores for test words with low _a priori_ intelligibility were in fact slightly higher than the test words with high _a priori_ intelligibility. This direction is contrary to that obtained for the deaf subjects.

Scores for test words in sentences did not differ as

TABLE 15

PERCENT INTELLIGIBILITY OF SCORES FOR HEARING CHILDREN AS A FUNCTION OF *A PRIORI* INTELLIGIBILITY AND CONTEXT

	High Context		Low Context	
	High Intell.	Low Intell.	High Intell.	Low Intell.
Test words in sentences	1.00	.98	.99	.96
Segmented test words	.82	.81	.57	.64

a function of sentence length, an average of about 96% being obtained for sentences. Scores for segmented test words showed some variation with the number of syllables. Scores for three and seven syllable sentences were approximately 73%, whereas scores for five syllable sentences were only 60%. These results differ from those obtained for deaf children in that for these children greater differences in intelligibility were obtained between 3, 5 and 7 syllable sentences while scores were very similar for segmented test words.

There were also differences between the scores for deaf and hearing children as a function of the position of the test word in the sentence. For hearing children, scores for test words near the beginning, middle, or end of the sentence were essentially the same, each averaging about 98%. For segmented test words, there was a small variation in test scores as a function of word position. Scores for segmented test words in the final position were slightly higher (73%) than those in the beginning or middle (69% and 63% respectively). The data obtained for the deaf children systematically decreased from beginning to middle to final positions for the test words in sentences, and were similar for all positions for segmented test words. The magnitude of the differences in scores although different in form are comparable between deaf and hearing children.

Finally, there were minor differences in the scores for the hearing children as a function of age or sex.

Summary of Results for Hearing Children

The data revealed several findings with respect to scores obtained for deaf and hearing children:

1. For both groups of children, scores for test words in sentences were higher than words in isolation, which in turn were higher than test words produced in sentences but heard in isolation (i.e. segmented test words). The differences in test scores as a function of the type of stimulus were much larger for the deaf children than the hearing children.

2. For hearing children, context had little effect on test scores for sentence material, these scores being essentially 100% for either high or low context sentences. For segmented test words, the scores for high context sentences were twenty percentage points higher than those of test words segmented from low context sentences. A context effect for segmented test words was also found for deaf children although the magnitude of this effect was smaller than for the hearing group.

3. For deaf children, there was a significant difference between test words with high and low *a priori* intelligibility for each type of stimulus. This was not the case for the hearing children.

4. For either test words in sentences or segmented test words, there were small negligible differences in test scores as a function of sentence length or word position. In contrast, scores obtained with the deaf children showed

evidence of an interaction between the number of syllables in the sentences and the word position.

CHAPTER V

DISCUSSION

Interpretation of the Results

The principle aim of this study was to investigate the differences between experienced and inexperienced listeners in understanding the speech of the deaf. The results showed quite clearly that intelligibility scores for the experienced listeners were consistently higher than those for the inexperienced listeners. Further, the differences in the test scores between experienced and inexperienced listeners were essentially constant for all factors investigated. These factors included: (1) the degree of sentence context, (2) the *a priori* intelligibility of a test word when produced by a deaf child, (3) the number of syllables in the sentence, and (4) the position of the test word in the sentence. It was also found that for both groups of listeners, the scores for test words in sentences were consistently higher than scores for segmented test words.

The results of this study are consistent with the findings reported in the literature. However, the data do not support several hypotheses that have been advanced to explain the differences between listeners.

One commonly held view is that the experienced listeners obtained higher scores than inexperienced listeners because they are familiar with typical errors in production of deaf speech, and recode the speech so as to compensate for these errors. If this were the case, one would expect the data to show a difference between experienced and inexperienced listeners in the relative gain of test scores for words with high or low _a priori_ intelligibility. By definition, words with high _a priori_ intelligibility were words which deaf children were likely to produce correctly (as determined empirically from Smith, 1972). Similarly, words with low _a priori_ intelligibility were words which deaf children were likely to misarticulate.

Hence, if the above hypothesis is correct, then experienced listeners would show a greater relative gain for low _a priori_ intelligibility words, since these should have more errors for the listener to recode. However, no significant interaction was observed between listener experience and _a priori_ intelligibility. The measured difference in scores between experienced and inexperienced listeners for test words with high _a priori_ intelligibility was about the same as those for test words with low _a priori_ intelligibility, as shown in Figure 8.

The lack of a statistically significant interaction between listener experience and _a priori_ intelligibility does not necessarily mean that experienced listeners recode deaf speech in the same way as inexperienced listeners as a function of relative intelligibility, but rather that should such an effect exist, it is more subtle and less easily detected than previously supposed.

A hypothetical example of the difficulty of detecting such a differential effect is as follows. Consider two samples of deaf speech, one with a 50% error rate in articulation, and the other with a 20% error rate. If the effect of listener experience in recoding the speech is equivalent to a 10% reduction in errors, then for the first sample, the error rate would be reduced to 45%, and for the second sample, changed to an error rate of 18%. An experiment of the type performed here would be quite likely to detect a significant difference in scores between experienced listeners, but there is a good chance that the differential improvement in test scores (between experienced and inexperienced listeners) as a function of the a priori intelligibility would be too small to be detected in an experiment of moderate size.

Another commonly held hypothesis, advanced by Hudgins and Numbers (1942) and also by Thomas (1963), is that experienced listeners simply make better use of contextual cues than inexperienced listeners, thereby increasing overall intelligibility. As is obvious from Figure 7, the scores for both experienced and inexperienced listeners were higher for sentences with high context, than for those with low context, but there was no evidence of a statistically significant interaction between listener experience and the degree of context. The improvement due to experience was essentially constant for both high context and low context stimuli. Again, the lack of a statistically significant interaction does not rule out the possibility of such an effect, but rather indicates that, should an interaction exist, it is likely to be of small magnitude.

While the importance of context has long been realized, it has been argued by Hudgins and Numbers (1942) that context may be even more important for listeners of deaf speech. Specifically, they hypothesized that the effect of articulatory errors on the intelligibility of deaf speech could be reduced by the contextual constraints of the sentences, and by implication, the greater the articulatory errors, the greater the effect of context.

This third hypothesis concerning an interaction between _a priori_ intelligibility and context, was supported by the data. The effect of _a priori_ intelligibility, from high to low, accounted for a greater change in scores for high context sentences than for low context sentences. While there was a significant interaction between _a priori_ intelligibility and context for test words in sentences, the interaction between these factors and listener experience was not statistically significant, suggesting that both experienced and inexperienced listeners are benefitting to the same extent from this information.

A fourth view, often held by persons not familiar with the deaf, is that personal knowledge of the deaf speaker enables the experienced listener to obtain higher intelligibility scores. Since the inexperienced listener does not know the speaker, his scores would be lower.

In the literature, a definition of experienced listener includes listeners who knew the children, such as the child's classroom teacher or the parents (Mangan, 1961), listeners who were trained on either the test materials or the deaf speakers (Hudgins, 1949), and also listeners who were generally familiar with the speech of the deaf, but did not personally know the speakers. In contrast, all inexperienced listeners were specified as having no previous experience with the deaf.

In this investigation, none of the listeners, experienced or inexperienced, knew the child whose speech they heard. Hence, the hypothesis of personal knowledge of the speaker alone enabling the experienced listener to obtain higher intelligibility scores was not supported by the study. While it is likely that children who are known to the listeners may be more intelligible to parents or classroom teachers than to other listeners, further research is warranted to quantify the effect of personal knowledge.

A final notion is that knowledge of a particular speech teaching strategy which results in a distinctive speech pattern, and which is characteristic of the child's school, enables the experienced listener to obtain higher intelligibility scores. Similarly, if other experienced listeners, or inexperienced listeners, are unfamiliar with this educational approach, the intelligibility scores will be lower.

This view is also not supported by the data. First, the error patterns of the children in this study were similar to other deaf children (Smith, 1972; Levitt et al., 1975). Second, in most cases the experienced listeners did not know the school at which the child received his training. Teachers who served as experienced listeners from the same school as the children, scored no better or worse than the experienced listeners from other schools. It would seem that once familiar with deaf speech, the experienced listeners were able to generate higher scores for deaf speakers in general.

The Effect of Context on Perception and Production

One can infer from the results of this study, that the effect of context is important in perception as well as in production. For the former, the effect of semantic context was seen in the differences in test scores for speech stimuli with high or low degrees of context, and also in the differences between test words produced and heard in sentences, and test words produced in sentences but heard in isolation (i.e. segmented). It should be remembered that the recordings of test words in sentences and in segmented conditions were identical.

The effect of phonetic context on production is noted in the differences in test scores between isolated words and segmented test words, the scores for the former being considerably higher. The difference in test scores indicates that deaf children produce words in context much differently than words in isolation. In other words, there are coarticulatory effects in the speech of the deaf which have a significant effect on intelligibility.

It would be wrong, however, to assume that since coarticulatory effects seem to have a negative effect on intelligibility, (as manifested by relatively low scores for segmented test words), that production of speech one-word-at-a-time would be helpful to improve intelligibility. While this study did not consider test words produced in isolation but heard in context, it is well known that speech produced by the concatenation of isolated words, without additional processing (Flanagan, 1972), is both difficult to understand and unpleasant to hear.

Another production effect observed was that the total energy for a word produced in isolation was different from that for the same word produced in sentences. Specifically, isolated test words tended to be more intense than those produced in sentences, and longer in duration. However, the perceptual differences observed between test words in sentences and in isolation cannot be ascribed to differences in intensity, since the levels for test words in each condition (sentences, isolation and segmented) were equalized.

Another important finding in production can be derived from the data for segmented test words. For the deaf children, there was a significant difference in test scores between words segmented from high context sentences, and those segmented from low context sentences, the former being more intelligible than the latter. For hearing children, the overall scores for segmented words from either high or low context sentences were nearly the same, with the latter test words slightly more intelligible. The direction of the context effect for the deaf children, and the similarity in scores for high or low context segmented test words for hearing children, is opposite that reported by Lieberman (1963).

Lieberman reported that scores for words segmented from redundant (i.e. high context) sentences were only half again as intelligible as those segmented from non-redundant (i.e. low context) sentences. Lieberman showed that words in low context sentences were produced with greater stress, longer duration or more precise articulation than high context words, thus helping the listener to resolve any linguistic ambiguities in the message.

There were, however, several important differences between this investigation and Lieberman's. First, Lieberman's subjects were adults, while those in this study were children. Second, the test words in the redundant and non-redundant sentences produced by the adults were the same words, while those produced by the children in this investigation were different words in the high context and low context sentences.

If, however, Lieberman's hypothesis were tenable, then the differences in scores between the children and the adults suggests that the younger group may not yet have learned to use different production strategies to assist the listener. Whether one can argue a trend in the data towards these strategies for hearing children, and their absence for the deaf children, requires further investigation.

Comparison with the Literature

While the results of this study do not support all of the hypotheses forwarded by previous investigators, the data obtained are not inconsistent with those of earlier studies. The principle study on the differences between experienced and inexperienced listeners to date was conducted by Thomas (1963). Scores obtained by Thomas and those of the present study are contrasted as follows:

Study	Isolated Words		Sentences	
	Exper.	Naive	Exper.	Naive
Thomas (Auditory Only)	37%	28%	77%	50%
Present Study	29%	23%	41%	30%

In both these investigations, scores for experienced listeners were higher, on the average, than those of inexperienced listeners, and the difference between listeners was greater for sentences than for isolated words. However, overall intelligibility scores in the present investigation were poorer than those reported by Thomas, expecially for sentences. There may be several reasons for the differences.

First, the audiological data were not reported for the deaf children in Thomas's study, while the subjects in the present study all had severe to profound hearing losses. Since several investigators (Boothroyd, 1974; Smith, 1972) have shown that there is a correlation between intelligibility of speech and hearing level (especially up to 90 dB), the differences in scores may be the result of differences in the children's hearing levels.

Secondly, the CID sentences of "everyday American speech" produced by the children in Thomas's study are noted for high contextual redundancy and common word usage (Davis and Silverman, 1978). In contrast, the speech stimuli in this study were controlled for the degree of context. However, even the data for high context sentences obtained in this investigation did not approach the scores reported by Thomas.

Since scores for isolated words were closely grouped, and those for sentences were more diverse, it may be that the differences between the investigations are due more to choice of sentence material than to differences in hearing levels.

As noted in the review of the literature, numerous investigators have measured the intelligibility of deaf speakers from scores obtained by experienced or inexperienced listeners. As previously noted, the definition of the respective groups of listeners was not always clearly specified.

For isolated words, the scores obtained by experienced listeners ranged form 35% (Subtelny, 1977) to 42% (Hudgins, 1949), with those of the experienced listeners in this study 29%. For inexperienced listeners, the scores ranged from 17% (Brannon, 1964) to 28% (Thomas, 1963), with scores for the inexperienced listeners in this study, 23%. It is interesting to note that when scores for test words are examined as a function of *a priori* intelligibility, those with high *a priori* intelligibility fell essentially mid-range of the published data for either experienced or inexperienced listeners. This suggests that the phonetically balanced monosyllables used in the earlier research may be more like test words with high *a priori* intelligibility, than like test words with low *a priori* intelligibility for deaf speakers, resulting in higher intelligibility scores.

Scores reported for sentences in the literature, cover a much greater range of intelligibility than scores for isolated words. For experienced listeners, scores ranged from 31% (Markides, 1970) to 83% (Monsen, 1978); for inexperienced listeners, the range was 18.7% (Smith, 1972) to 73% (Monsen, 1978). Scores for test words in sentences in this study were 41% for experienced, and 30% for inexperienced listeners, with scores for all words in sentences, only slightly higher (49% and 35%, respectively).

The scores for test words with a high degree of context (49% for experienced and 38% for inexperienced listeners) were mid-range of those reported in the literature.

Scores for test words with a low degree of context (33% for experienced, and 21% for inexperienced listeners), fell near the lower end of the range for the respective groups.

Apart from the present study which controlled for the degree of context, the speech materials which resulted in lower scores were either spontaneous speech samples (John and Howarth, 1964; Markides, 1970) or sentences which varied considerably in length and grammatical complexity (Smith, 1972). Speech materials that resulted in high intelligibility were those which contained words of common usage or were highly redundant (Thomas, 1963; Monsen, 1978). The wide variations in scores reported for deaf children with similiar hearing losses, implies the necessity for a set of uniform speech materials, thus permitting more meaningful evaluation of intelligibility, and comparison between deaf speakers.

Implications

Of the variables considered in this study, only the stimulus type (test words in sentences, in isolation or in segmented conditions) showed any evidence of a possible interaction with listener experience. That is, the difference between experienced and inexperienced listeners was greater in sentences than in isolation.

The finding of no significant interaction between listener experience and any of the other factors investigated implies that the effect of experience is not due to any superficial recoding of deaf speech on the part of the experienced listeners. If the factors considered in this study (i.e. context, *a priori* intelligibility, sentence length or word position) were the keys to the differences between listeners, then marked improvement in the intelligibility of deaf speech for the "man on the street" could be accomplished by a training program that concentrated on those factors most responsible for the differences between listeners. Training procedures are not that simple, and the underlying mechanism for why experienced listeners do so much better than inexperienced listeners is not yet known.

In addition to the main effects tested, it is also known that the difference between experienced and inexperienced listeners are not due to any secondary effects such as idiosyncracies in particular children, in specific test words, or in either groups of listeners. As shown in Figures 11-13, the scores obtained by experienced or inexperienced listeners for individual children in the study were roughly parallel, regardless of relative intelligibility. Any variation from the pattern was easily explained by data obtained from a second set of listeners. Overall scores for younger children were slightly poorer than those for older children, as was also observed by Smith (1972). However, there were essentially no differences between male and female speakers. Similarly, examining the scores obtained by experienced and inexperienced listeners for individual test words did not reveal any unusual variation from the patterns obtained for any other variables in the study.

Finally, the data revealed that experienced and inexperienced listeners showed no difference in the relative variability of the scores for either test words in sentences or isolated words. An exception for segmented words showed a slightly larger variability for inexperienced listeners. However, examination of the data showed that on the average, scores were low for inexperienced listeners in this condition, and scores of zero, which inflated the estimate of the variance, occurred more frequently.

In sum, the difference between experienced and inexperienced listeners cannot be accounted for in any obvious way. For each factor, analysis of the data indicates a remarkably constant difference between groups. The result of this finding suggests that the advantage of experience cannot be attributed simply to one or two variables, at least within the factors considered within this study.

Consequently, the differences between experienced and inexperienced listeners must be due to fairly complex aspects of deaf speech which are not immediately apparent to the listener, but which must be learned. The fact that the difference is constant indicates that the effect occurs consistently over a wide range of variables and there is a need for additional research. Such research might include studies of: the effect of the personal knowledge of the speaker; the importance of visual cues; how spectral information in the speech of the deaf is coded differently from that of normals; and how coarticulatory phenomena are manifested in the speech of the deaf.

CHAPTER VI

SUMMARY

The study was undertaken in order to quantify the differences between experienced and inexperienced listeners in understanding the speech of the deaf. The specific questions asked were as follows:

1. Is there a significant difference between experienced and inexperienced listeners?

2. What is the effect of listener experience on the intelligibility of words produced in sentences, in isolation, and produced in sentences but heard in isolation, (having been segmented from the sentence)?

3. What is the effect between listener experience and context, e.g. does context benefit the experienced listeners more than the inexperienced listeners?

4. Are any of the above effects influenced by the relative intelligibility of the test word?

One hundred twenty listeners, sixty listeners highly experienced in hearing the speech of the deaf, and sixty listeners with no previous experience in hearing the speech of the deaf, participated in the study. An additional twenty-four listeners heard the tapes produced by a control group of hearing children.

The speech stimuli consisted of a list of 36 test words in sentences, and a list of the same 36 test words produced in isolation. These stimuli were recorded by twenty deaf children as well as by eight hearing children who served as the control group.

The speech stimuli were designed to include the following factors: (1) an estimate of the _a priori_ intelligibility of a test word when produced by a deaf child (i.e. high _a priori_ intelligibility versus low _a priori_ intelligibility), (2) an estimate of the degree of sentence context (i.e. high sentence context versus low sentence context), (3) sentences that varied in length (i.e. three, five or seven syllable sentences), and (4) test words that varied in the position of the sentence (i.e. near the beginning, the middle, or the end of the sentence). The 36 test words in sentences included all of these factors; the 36 test words in isolation included only _a priori_ intelligibility.

Since factors in both production and perception may account for any measured difference between test scores for words produced in sentences and in isolation, an additional experimental condition was included. Specifically, this was test words produced in sentences but heard in isolation having been segmented out of the sentence using computer techniques.

Since the intensity levels of each test word varied from sentences, to isolated, to segmented conditions, graphic level recordings were made for the speech stimuli, and the levels of each word produced by a single child were adjusted to within ± 1 dB of the reference level which was that of the test word in the sentence.

Thus, each listener was presented with three types of speech stimuli. These were:

1. Test words produced in sentences and presented to the listener in sentences.

2. Test words produced in isolation and presented to the listener in isolation.

3. Test words produced in sentences, segmented from the sentences, and presented to the listener in isolation.

These three conditions are referred to as test words in sentences, test words in isolation, and segmented test words, respectively.

A listener heard 36 items produced by a single deaf or a single hearing child, twelve items for each of the above three experimental conditions. In order to avoid learning effects, there was no repetition of a test word on any tape, and each listener heard only one tape.

The data revealed several findings with respect to the differences between experienced and inexperienced listeners. Scores for the experienced listeners were consistently higher than those for the inexperienced listeners for all conditions. The magnitude of the differences between listeners was largest for test words in sentences, about two-thirds as large for test words in isolation, and roughly half again as large for segmented test words.

Further, there were no statistically significant interactions between listener experience and the other factors considered in this study, specifically the degree of sentence context, the *a priori* intelligibility of the test words, the sentence length, or the position of the test word in the sentence. It was also found that the difference attributed to experience was not due to any idiosyncracies in particular children, in specific test words, or within either group of listeners.

In addition, for both groups of listeners the pattern of intelligibility scores were similar. These are briefly summarized as follows. Scores for test words with high *a priori* intelligibility were higher than those with low *a priori* intelligibility. Also test words in high context sentences were more intelligible than those in low context sentences. Intelligibility scores did not increase as the overall length of the sentence increased. Scores for test words in the word initial position of three syllable sentences were consistently lower than for other locations in the sentence. However, for five and seven syllable sentences, the reverse effect was observed, words produced near the beginning of a sentence having higher intelligibility scores.

Finally, the scores for older children were slightly higher than those of younger children, but there was no significant difference between male and female speakers.

For hearing children, the ranking of the relative intelligibility across experimental conditions (when possible) was similar to that obtained for the deaf children, although the scores were much higher on average. Also, the magnitude of the differences between conditions were quite small, and for some factors a ranking was not possible since many of the scores were either at or approached 100%.

The results of this study are consistent with the findings reported in the literature; namely, that experienced listeners obtain higher intelligibility scores than inexperienced listeners. However, the data do not support several hypotheses that have been advanced to explain why the experienced listeners obtain higher scores.

The effect of context was found to be important in perception as well as in production. For the former, the effect of semantic context was observed in the differences in test scores for speech stimuli with high or low degrees of context, and also in the differences between test words produced in sentences and test words produced in sentences but heard in isolation (i.e. segmented).

The effect of phonetic context on production was noted in the differences in scores between isolated test words and segmented test words, the scores for the former being considerably higher. This indicates that the deaf child produces words in context differently than words in isolation. That is, there are coarticulatory effects in the speech of the deaf which have a significant effect on intelligibility.

The findings of this study imply that the effect of experience is not due to any superficial recoding of deaf speech on the part of the listeners, at least within the factors considered here. The differences between groups of listeners were remarkably constant. Consequently, the differences between experienced and inexperienced listeners must be due to fairly complex aspects of deaf speech which are not immediately apparent to the listener and must be learned. Further research is necessary to delineate the reasons for the differences between experienced and inexperienced listeners.

APPENDICES

APPENDIX A

PURE TONE AVERAGES OF EACH DEAF CHILD IN dB RE. AUDIOMETRIC ZERO (ISO)

	Young Deaf Children				Old Deaf Children			
	Males		Females		Males		Females	
	Right	Left	Right	Left	Right	Left	Right	Left
1	98	100	110+	107+	103	103	102	99
2	95	78	110+	82+	105	100	103	108
3	106	110	110+	110+	98	98	97	105
4	100	110	105+	105+	102	105	109	107
5	110	107	102+	85	97	92	93	95

APPENDIX B

Procedure for the Specification of the Speech Levels

The procedure for specifying the speech levels was accomplished in two parts: (1) the intensity levels of the test words in sentences, in isolation and segmented were measured, and (2) the intensity levels were equalized so that each test word was presented to the listener at the same level regardless of the experimental condition.

Graphic level recordings were made of the tapes of the test words in sentences, in isolation and segmented conditions. The measurements were obtained using a Bruel and Kjaer Graphic Level Recorder (Model 2305).

Prior to actual measurements, level recordings of selected samples of the deaf children's speech were obtained at different settings of the instrument. Of particular concern was minimizing the amount of overshoot and undershoot, since many of the test words produced by the deaf children varied greatly in duration and intensity. Overshoot is evidenced by an inordinate deflection of the pen, which results in inaccurate high measurements of the input signal. A problem of the opposite nature would be undershoot. In this case, the pen deflection does not reach the target value of the input signal, and the measurements are inaccurately low.

These problems were resolved in two ways. First, the writing speed of the level recorder seemed to contribute to changes observed in preliminary measurements. Therefore this setting was systematically increased and decreased until evidence of overshoot or undershoot in the children's speech production samples occurred. A setting was then chosen below or above the writing speed which produced the inaccurate measurements.

Secondly, it was also found that the effective response time of the level recorder varied with magnitude of the pen deflection. The tape replay system and the input attenuator of the level recorder were therefore adjusted in an effort to obtain a maximum pen excursion of 40 mm. This was not always possible since many of the deaf children produced the speech samples with wide fluctuations in loudness.

The level recordings of the deaf children's speech were obtained with the instrument set to operate as follows:

1. Potentiometer Range - 50 dB
2. Rectifier Response - Peak
3. Lower Limiting Frequency - 50 Hz
4. Writing Speed - 400 mm/sec
5. Paper Speed - 30 mm/sec
6. Input potentiometer - 0 dB
7. Input Attenuator - variable

For level recordings of the hearing children's speech, the instrument settings were the same as the deaf children's, with the exception of the writing speed set at 250 mm/sec.

After graphic level recordings of the original tapes had been obtained, measurements were made of each test word in a sentence, in isolation, and segmented. The intensity level (in decibels) measured for each test word in a sentence was set as the reference level. The differences (in decibels) between the reference levels (for each individual test word) and the intensity levels for the test word in isolation and segmented were noted. These numbers obtained for isolated and segmented test words represented the amount of attenuation or amplification necessary to equalize the test word in each listening condition. That is, the same test word heard in sentences, isolation or segmented conditions would be presented at an equal intensity level for each test word in the isolated and segmented conditions would be presented at an equal intensity level $\pm$ 1 dB.

In the second part of the procedure, the intensity level for each test word in the isolated and segmented conditions was adjusted. A block diagram of the equipment used to obtain equalized recordings of the speech stimuli is shown in Figure 24.

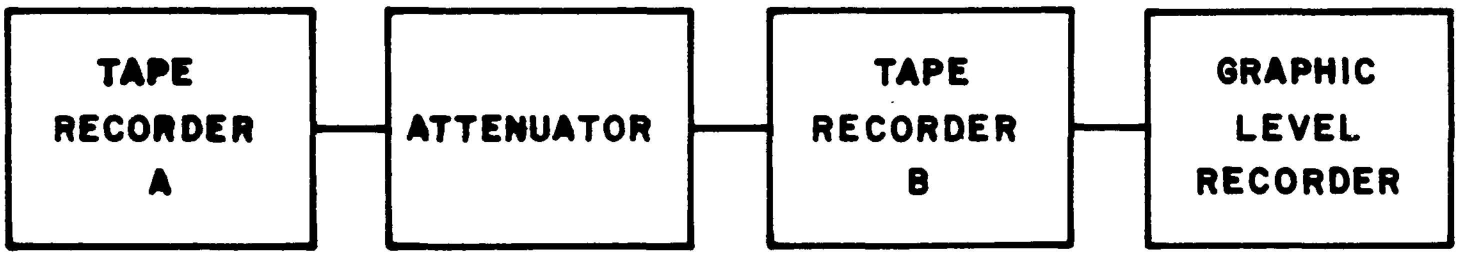
The original tape recordings of test words in isolation or segmented test words were first placed on tape recorder B (Ampex Model 440B). A second set of level recordings were then obtained for several words on the tape using the same level recording settings as in the original measurement procedure. These second level recordings were compared to the original measurements and served solely as a check to insure that all settings on the graphic level recorder and tape recorder B were replicated exactly.

The original tape placed on tape recorder A (Ampex AG-500) was then re-recorded on tape recorder B. The intervening attenuators (Daven Models 2513 and 2511), together with the fixed gain of the tape playback system, permitted the speech to be attenuated or amplified in a combination of 10 dB and 1 dB steps respectively. These attenuator settings were systematically adjusted in turn so that the equalized version of the isolated or segmented test word would be at the level of the test word in the sentence $\pm$ 1 dB.

Level recordings were obtained immediately after re-recording to insure that each test word was equalized correctly. This procedure was repeated for each deaf and each hearing child's recordings of test words in isolation and segmented test words until equalized versions of all original materials, except the reference test words in sentences, were obtained.

Following the equalizing procedure, a 400 Hz calibration tone was generated which preceded each child's 3 submaster tapes: test words in sentences, test words in isolation, and segmented test words. The level of the calibration tone was the mean of the 36 test words in sentences for each child. This calibration tone was used to maintain the equalized levels during the randomization procedure and during playback.

Figure 24. The block diagram of the equipment used to equalize the speech stimuli.


TAPE
RECORDER
A
ATTENUATOR
TAPE
RECORDER
B
GRAPHIC
LEVEL
RECORDER

APPENDIX C

SEQUENCE OF THE TEST ITEMS AND FACTOR ANALYSIS

TAPE 1

	Test Item	Stimulus	A Priori Intell.	Context	Number of Syllables	Position in Sentence
Test words in sentences	1	If it's cool I cannot go.	Low	Low	7	Medial
	2	Did you brush your teeth?	Low	High	5	Final
	3	Feed the dog.	Low	Low	3	Begin.
	4	The flag is red, white and blue.	High	High	7	Medial
	5	He said he could go.	Low	Low	5	Final
	6	Is there no more milk?	High	High	5	Final
	7	Read the book.	High	High	3	Begin.
	8	Have a lot.	High	Low	3	Begin.
	9	We have food for the picnic.	Low	High	7	Medial
	10	It's easy to hear her.	High	Low	5	Final
	11	Keep quiet.	Low	High	3	Begin.
	12	Is the fat baby crying?	High	Low	7	Medial
Segmented test words	13	will	High			
	14	did	Low			
	15	dog	Low			
	16	that	Low			
	17	piece	High			
	18	cat	Low			
	19	fall	High			
	20	name	High			
	21	with	High			
	22	beach	Low			
	23	need	High			
	24	deep	Low			

APPENDIX C-Continued

	Test Item	Stimulus	A Priori Intell.	Context	Number of Syllables	Position in Sentence
Isolated test words	25	this	High			
	26	was	Low			
	27	good	High			
	28	man	High			
	29	hair	Low			
	30	coat	High			
	31	cake	High			
	32	wish	Low			
	33	has	Low			
	34	his	Low			
	35	ball	Low			
	36	book	High			

APPENDIX C-Continued

TAPE 2

	Test Item	Stimulus	A Priori Intell.	Context	Number of Syllables	Position in Sentence
Segmented test words	1	cake	High			
	2	this	High			
	3	wish	Low			
	4	ball	Low			
	5	his	Low			
	6	man	High			
	7	has	Low			
	8	good	High			
	9	book	High			
	10	coat	High			
	11	was	Low			
	12	hair	Low			
Isolated test words	13	more	High			
	14	could	Low			
	15	red	High			
	16	teeth	Low			
	17	have	High			
	18	keep	Low			
	19	hear	High			
	20	feed	Low			
	21	fat	High			
	22	read	High			
	23	food	Low			
	24	cool	Low			

APPENDIX C-Continued

	Test Item	Stimulus	A Priori Intell.	Context	Number of Syllables	Position in Sentence
Test words in sentences	25	It is nice on a fall day.	High	Low	7	Final
	26	The dog barks.	Low	High	3	Medial
	27	My name is Nancy.	High	High	5	Begin.
	28	Come with me.	High	High	3	Medial
	29	We will go to the beach today.	Low	Low	7	Final
	30	I need it.	High	Low	3	Medial
	31	Can you dive in deep water?	Low	High	7	Final
	32	Is that the tall one?	Low	Low	5	Begin.
	33	You did it.	Low	Low	3	Medial
	34	They will come again.	High	Low	5	Begin.
	35	May I have a piece of cake.	High	High	7	Final
	36	The cat chased the mouse.	Low	High	5	Begin.

APPENDIX C-Continued

TAPE 3

	Test Item	Stimulus	A Priori Intell.	Context	Number of Syllables	Position in Sentence
Isolated test words	1	name	High			
	2	need	High			
	3	deep	Low			
	4	cat	Low			
	5	dog	Low			
	6	that	Low			
	7	beach	Low			
	8	with	High			
	9	fall	High			
	10	did	Low			
	11	piece	High			
	12	will	High			
Test words in sentences	13	Who wants this ice cream?	High	Low	5	Medial
	14	I wish I had a pony.	Low	High	7	Begin.
	15	Comb your hair.	Low	High	3	Final
	16	Get your ball and bat.	Low	High	5	Medial
	17	What was the name of that boy?	Low	Low	7	Begin.
	18	Get the cake.	High	Low	3	Final
	19	That's no good.	High	High	3	Final
	20	Get your coat and hat.	High	High	5	Medial
	21	Mother has the car.	Low	Low	5	Medial
	22	The book is on the table.	High	Low	7	Begin.
	23	This is his.	Low	Low	3	Final
	24	That man is not my father.	High	High	7	Begin.

APPENDIX C-Continued

	Test Item	Stimulus	A Priori Intell.	Context	Number of Syllables	Position in Sentence
Segmented test words	25	food	Low			
	26	hear	High			
	27	keep	Low			
	28	cool	Low			
	29	have	High			
	30	more	High			
	31	read	High			
	32	could	Low			
	33	fat	High			
	34	red	High			
	35	feed	Low			
	36	teeth	Low			

APPENDIX D

Instructions to Listeners of the Speech of the Deaf

You are going to hear 36 utterances recorded by a deaf child. Twelve of the utterances are sentences; 24 of them are single words. There are no repetitions of words or sentences. Each utterance is different. Each utterance will be played twice.

You are to write down whatever you think the child is saying. Note: even the best of these children do not sound like normal hearing speakers. If some words are unclear, you may guess. If you can catch only isolated words in the sentence, write them down, indicating their approximate position in the sentence. An example of this is:

_______ _____ boy _____ not ______.

Have you any questions?

APPENDIX D

Instructions to Listeners of Normal Speech

You are going to hear 36 utterances recorded by a normal child. Twelve of the utterances are sentences; 24 of them are single words. There are no repetitions of words or sentences. Each utterance is different. Each utterance will be played twice.

You are to write down whatever you think the child is saying. If some words are unclear, you may guess. If you can catch only isolated words in the sentence, write them down, indicating their approximate position in the sentence. An example of this is: _______ _____ boy ______ not _______.

Have you any questions?

APPENDIX E

A Listing of the Intelligibility Scores Obtained by Experienced and Inexperienced Listeners for Each Deaf Speaker

SUBJECT 1

	EXPERIENCED 1	2	3	AVERAGE	NAIVE 61	62	63	AVERAGE	OVERALL AVERAGE
SENTENCES	0.167	0.333	0.417	0.306	0.167	0.167	0.167	0.167	0.236
COMPUTER WORDS	0.167	0.167	0.250	0.194	0.0	0.250	0.250	0.167	0.181
ISOLATED WORDS	0.417	0.333	0.083	0.278	0.167	0.250	0.083	0.167	0.222

SUBJECT 2

	EXPERIENCED 4	5	6	AVERAGE	NAIVE 64	65	66	AVERAGE	OVERALL AVERAGE
SENTENCES	0.083	0.417	0.167	0.222	0.167	0.167	0.083	0.139	0.181
COMPUTER WORDS	0.083	0.167	0.167	0.139	0.083	0.083	0.250	0.139	0.139
ISOLATED WORDS	0.417	0.250	0.083	0.250	0.250	0.167	0.167	0.194	0.222

SUBJECT 3

	EXPERIENCED 7	8	9	AVERAGE	NAIVE 67	68	69	AVERAGE	OVERALL AVERAGE
SENTENCES	0.250	0.500	0.083	0.278	0.083	0.250	0.167	0.167	0.222
COMPUTER WORDS	0.167	0.167	0.0	0.111	0.083	0.0	0.250	0.111	0.111
ISOLATED WORDS	0.167	0.083	0.250	0.167	0.167	0.167	0.167	0.167	0.167

SUBJECT 4

	EXPERIENCED 10	11	12	AVERAGE	NAIVE 70	71	72	AVERAGE	OVERALL AVERAGE
SENTENCES	0.333	0.500	0.583	0.472	0.250	0.250	0.167	0.222	0.347
COMPUTER WORDS	0.250	0.333	0.167	0.250	0.083	0.250	0.083	0.139	0.194
ISOLATED WORDS	0.583	0.333	0.333	0.417	0.333	0.250	0.0	0.194	0.306

SUBJECT 5

	EXPERIENCED 13	14	15	AVERAGE	NAIVE 73	74	75	AVERAGE	OVERALL AVERAGE
SENTENCES	0.083	0.333	0.083	0.167	0.0	0.250	0.250	0.167	0.167
COMPUTER WORDS	0.083	0.083	0.0	0.056	0.083	0.083	0.0	0.056	0.056
ISOLATED WORDS	0.250	0.0	0.0	0.083	0.0	0.0	0.0	0.0	0.042

SUBJECT 6

	EXPERIENCED				NAIVE				
	16	17	18	AVERAGE	76	77	78	AVERAGE	OVERALL AVERAGE
SENTENCES	0.250	0.333	0.417	0.333	0.333	0.333	0.250	0.306	0.319
COMPUTER WORDS	0.167	0.417	0.083	0.222	0.167	0.167	0.0	0.111	0.167
ISOLATED WORDS	0.167	0.250	0.250	0.222	0.167	0.167	0.0	0.111	0.167

SUBJECT 7

	EXPERIENCED				NAIVE				
	19	20	21	AVERAGE	79	80	81	AVERAGE	OVERALL AVERAGE
SENTENCES	0.667	0.833	0.917	0.806	0.667	0.750	0.750	0.722	0.764
COMPUTER WORDS	0.250	0.500	0.167	0.306	0.167	0.167	0.167	0.167	0.236
ISOLATED WORDS	0.750	0.583	0.583	0.639	0.667	0.333	0.583	0.528	0.583

SUBJECT 8

	EXPERIENCED				NAIVE				
	22	23	24	AVERAGE	82	83	84	AVERAGE	OVERALL AVERAGE
SENTENCES	0.583	0.750	0.750	0.694	0.750	0.583	0.750	0.694	0.694
COMPUTER WORDS	0.333	0.417	0.417	0.389	0.167	0.250	0.500	0.306	0.347
ISOLATED WORDS	0.667	0.583	0.333	0.528	0.500	0.417	0.417	0.444	0.486

SUBJECT 9

	EXPERIENCED				NAIVE				
	25	26	27	AVERAGE	85	86	87	AVERAGE	OVERALL AVERAGE
SENTENCES	0.083	0.417	0.500	0.333	0.083	0.250	0.250	0.194	0.264
COMPUTER WORDS	0.083	0.083	0.167	0.111	0.0	0.167	0.167	0.111	0.111
ISOLATED WORDS	0.417	0.167	0.167	0.250	0.167	0.083	0.167	0.139	0.194

SUBJECT 10

	EXPERIENCED				NAIVE				
	28	29	30	AVERAGE	88	89	90	AVERAGE	OVERALL AVERAGE
SENTENCES	0.500	0.500	0.833	0.611	0.750	0.583	0.667	0.667	0.639
COMPUTER WORDS	0.167	0.417	0.083	0.222	0.083	0.417	0.417	0.306	0.264
ISOLATED WORDS	0.250	0.333	0.500	0.361	0.250	0.417	0.500	0.389	0.375

SUBJECT 11

	Experienced				Naive				
	31	32	33	Average	91	92	93	Average	Overall Average
SENTENCES	0.333	0.667	0.417	0.472	0.167	0.417	0.333	0.306	0.389
COMPUTER WORDS	0.250	0.250	0.0	0.167	0.083	0.167	0.0	0.083	0.125
ISOLATED WORDS	0.333	0.250	0.250	0.278	0.333	0.083	0.250	0.222	0.250

SUBJECT 12

	Experienced				Naive				
	34	35	36	Average	94	95	96	Average	Overall Average
SENTENCES	0.167	0.417	0.250	0.278	0.0	0.167	0.083	0.083	0.181
COMPUTER WORDS	0.083	0.083	0.083	0.083	0.083	0.0	0.167	0.083	0.083
ISOLATED WORDS	0.333	0.333	0.167	0.278	0.417	0.167	0.083	0.222	0.250

SUBJECT 13

	Experienced				Naive				
	37	38	39	Average	97	98	99	Average	Overall Average
SENTENCES	0.667	0.667	0.750	0.694	0.167	0.333	0.417	0.306	0.500
COMPUTER WORDS	0.083	0.083	0.167	0.111	0.083	0.0	0.083	0.056	0.083
ISOLATED WORDS	0.417	0.333	0.167	0.306	0.667	0.250	0.250	0.389	0.347

SUBJECT 14

	Experienced				Naive				
	40	41	42	Average	100	101	102	Average	Overall Average
SENTENCES	0.417	0.583	0.583	0.528	0.250	0.500	0.417	0.389	0.458
COMPUTER WORDS	0.167	0.250	0.083	0.167	0.167	0.083	0.167	0.139	0.153
ISOLATED WORDS	0.667	0.333	0.167	0.389	0.500	0.333	0.167	0.333	0.361

SUBJECT 15

	Experienced				Naive				
	43	44	45	Average	103	104	105	Average	Overall Average
SENTENCES	0.417	0.583	0.667	0.556	0.250	0.417	0.667	0.444	0.500
COMPUTER WORDS	0.250	0.250	0.083	0.194	0.250	0.083	0.083	0.139	0.167
ISOLATED WORDS	0.667	0.417	0.167	0.417	0.417	0.250	0.333	0.333	0.375

SUBJECT 16

	EXPERIENCED				NAIVE				
	46	47	48	AVERAGE	106	107	108	AVERAGE	OVERALL AVERAGE
SENTENCES	0.083	0.250	0.250	0.194	0.0	0.250	0.0	0.083	0.139
COMPUTER WORDS	0.083	0.0	0.0	0.028	0.167	0.0	0.083	0.083	0.056
ISOLATED WORDS	0.083	0.083	0.083	0.083	0.083	0.250	0.083	0.139	0.111

SUBJECT 17

	EXPERIENCED				NAIVE				
	49	50	51	AVERAGE	109	110	111	AVERAGE	OVERALL AVERAGE
SENTENCES	0.167	0.667	0.667	0.500	0.167	0.417	0.667	0.417	0.458
COMPUTER WORDS	0.167	0.250	0.250	0.222	0.083	0.083	0.167	0.111	0.167
ISOLATED WORDS	0.250	0.250	0.417	0.306	0.250	0.083	0.333	0.222	0.264

SUBJECT 18

	EXPERIENCED				NAIVE				
	52	53	54	AVERAGE	112	113	114	AVERAGE	OVERALL AVERAGE
SENTENCES	0.167	0.333	0.167	0.222	0.167	0.250	0.167	0.194	0.208
COMPUTER WORDS	0.083	0.167	0.083	0.111	0.0	0.333	0.0	0.111	0.111
ISOLATED WORDS	0.167	0.167	0.083	0.139	0.250	0.167	0.083	0.167	0.153

SUBJECT 19

	EXPERIENCED				NAIVE				
	55	56	57	AVERAGE	115	116	117	AVERAGE	OVERALL AVERAGE
SENTENCES	0.250	0.250	0.083	0.194	0.083	0.083	0.083	0.083	0.139
COMPUTER WORDS	0.0	0.0	0.083	0.028	0.0	0.0	0.0	0.0	0.014
ISOLATED WORDS	0.250	0.083	0.167	0.167	0.167	0.083	0.083	0.111	0.139

SUBJECT 20

	EXPERIENCED				NAIVE				
	58	59	60	AVERAGE	118	119	120	AVERAGE	OVERALL AVERAGE
SENTENCES	0.250	0.333	0.250	0.278	0.083	0.167	0.333	0.194	0.236
COMPUTER WORDS	0.0	0.250	0.083	0.111	0.0	0.083	0.167	0.083	0.097
ISOLATED WORDS	0.333	0.250	0.0	0.194	0.083	0.083	0.0	0.056	0.125

APPENDIX F

A Listing of the Intelligibility Scores for Hearing Children

SUBJECT 30

	150	151	152	AVERAGE
SENTENCES	0.917	1.000	1.000	0.972
COMPUTER WORDS	0.667	0.417	0.750	0.611
ISOLATED WORDS	0.833	0.583	1.000	0.806

SUBJECT 31

	153	154	155	AVERAGE
SENTENCES	1.000	1.000	0.917	0.972
COMPUTER WORDS	0.500	0.750	0.500	0.583
ISOLATED WORDS	1.000	0.917	1.000	0.972

SUBJECT 32

	156	157	158	AVERAGE
SENTENCES	1.000	1.000	1.000	1.000
COMPUTER WORDS	0.583	0.583	0.917	0.694
ISOLATED WORDS	0.917	0.917	0.833	0.889

SUBJECT 33

	159	160	161	AVERAGE
SENTENCES	0.833	1.000	1.000	0.944
COMPUTER WORDS	0.750	0.500	0.833	0.694
ISOLATED WORDS	1.000	1.000	1.000	1.000

SUBJECT 34

	162	163	164	AVERAGE
SENTENCES	1.000	1.000	1.000	1.000
COMPUTER WORDS	0.833	0.583	1.000	0.806
ISOLATED WORDS	1.000	1.000	0.917	0.972

SUBJECT 35

	165	166	167	AVERAGE
SENTENCES	1.000	1.000	1.000	1.000
COMPUTER WORDS	0.917	0.583	0.750	0.750
ISOLATED WORDS	1.000	0.833	1.000	0.944

SUBJECT 36

	168	169	170	AVERAGE
SENTENCES	1.000	0.917	1.000	0.972
COMPUTER WORDS	0.667	0.667	0.833	0.722
ISOLATED WORDS	1.000	0.917	1.000	0.972

SUBJECT 37

	171	172	173	AVERAGE
SENTENCES	1.000	1.000	1.000	1.000
COMPUTER WORDS	0.583	0.917	0.917	0.806
ISOLATED WORDS	0.917	0.667	0.750	0.778

REFERENCES

Angelocci, A., Kopp, G., and Holbrook, A., The vowel formants of deaf and normal hearing eleven to fourteen year old boys. Jour. Speech and Hearing Dis., 29, 1964, 156-170.

Boothroyd, A., In, R. Stark (Ed.), Sensory Capabilities of Hearing-Impaired Children. Baltimore: University Park Press, 1974.

Brannon, J.B., Visual feedback of glossal motions and its influence on the speech of deaf children. Ph.D. Thesis, Northwestern Univ., 1964.

Brownlee, K.A., Statistical Theory and Methodology in Science and Engineering. New York: John Wiley and Sons, 1965.

Calvert, D.R., Some acoustic characteristics of the speech of profoundly deaf individuals. Ph.D. Thesis, Stanford Univ., 1961.

Davis, H. and Silverman, S.R., Hearing and Deafness (4th edition). New York: Holt, Rinehart and Winston, 1978.

Egan, J. P., Articulation testing methods, Laryng., 58, 1948, 955-991.

Flanigan, J. L., Speech Analysis Synthesis and Perception. New York: Springer-Verlag, 1972.

Green, D.S., Fundamental frequency of the speech of profoundly deaf individuals. Ph.D. Thesis, Purdue Univ., 1956.

Hood, R. B., Some physical concomitants of the perception of speech rhythm of the deaf. Ph.D. Thesis, Stanford Univ., 1966.

Hudgins, C. V., A method of appraising the speech of the deaf, Volta Review, 51, 1949, 597-638.

Hudgins, C. V., and Numbers, F. C., An investigation of the intelligibility of the speech of the deaf, Genet. Psych. Monog., 25, 1942, 289-392.

John, J. E. J., and Howarth, J. N., The effect of time distortions on the intelligibility of deaf children's speech. Lang. and Speech, 8, 1965, 127-134.

Klatt, D., The linguistic uses of segmental duration in English: acoustic and perceptual evidence. Paper presented at the 1975 Annual Convention of the American Speech and Hearing Association, Washington, D.C., 1975.

Kuhn, G. M., User's guide to SEDIT, Haskins Laboratories, 1974, (mimeographed)

Kuhn, G. M., User's guide to CONVERT, Haskins Laboratories, 1975, (mimeographed)

Levitt, H., Smith, C. R., and Stromberg, H., Acoustic, articulatory, and perceptual characteristics of the speech of deaf children. In, Proc. of the Speech Communication Seminar, Stockholm, G. Fant. (Ed.), New York: John Wiley and Sons, 1974, 129-139.

Levitt, H., Stark, R., McGarr, N.S., Carp, J., Stromberg, H., Gaffney, R. S., Barry, C., Velez, A., Osberger, M. J., Leiter, E., and Freeman, L., Language and communication skills of deaf children, 1973-1975, unpublished report.

Lieberman, P., Some effects of semantic and grammatical context on the production and perception of speech. Lang. and Speech, 6, 1963, 172-187.

McGarr, N. S., and Osberger, M. J., Pitch deviation and the intelligibility of deaf children's speech. Jour. of Comm. Disorders, 11, 1978, 237-247.

Mangan, K., Speech improvement through articulation testing. Amer. Annals of Deaf, 106, 1961, 39]-396.

Markides, A., The speech of deaf and partially hearing children with special reference to factors affecting intelligibility. Brit. Jour. Disorders of Communication, 5, 1970, 126-140.

Miller, G. A., Heise, G. A., and Lichten, W., The intelligibility of speech as a function of the context of the test materials. Jour. Exp. Psych., 41, 1951, 329-335.

Milner, P., Advantages of experienced listeners in intelligibility testing. IEEE Transactions on Audio and Electroacoustics, 1973, 161-165.

Monsen, R. B., Durational aspects of vowel production in the speech of deaf children. Jour. Speech and Hearing Res., 17, 1974, 386-398.

Monsen, R. B., Toward measuring how well hearing impaired children speak. Jour. Speech and Hearing Res., 21, 1978, 197-219.

Moser, H. and Dreher, J., Effect of training on listeners in intelligibility studies. Jour. Acoust. Soc. of Amer., 27, 1955, 1213-1219.

Nickerson, R. S., Computerized speech-training aids for the deaf. Report No. 2366, Bolt, Beranck and Newman, Inc. September 1973.

Nickerson, R. S., Stevens, K. N., Boothroyd, A., and Rollins, A., Some observations on timing in the speech of deaf and hearing speakers. Report No. 2905, Bolt, Beranck and Newman, Inc., September 1974.

Oller, D. K., The effect of position in the utterance on speech segment duration in English. Jour. Acoust. Soc. Amer., 54, 1973, 1235-1247.

O'Neill, J., Recognition of intelligibility test materials in context and isolation. Jour. Speech and Hearing Dis., 22, 1957, 87-90.

Osberger, M. J., The effect of timing errors on the intelligibility of deaf children's speech. Ph.D. Thesis, City University of New York, 1978.

Parkhurst, B. and Levitt, H., The effect of selected prosodic errors on the intelligibility of deaf speech. Jour. Comm. Disorders, 11, 1978, 249-256.

Pickett, J. M. and Pollack, I., Intelligibility of excerpts from fluent speech: effects of rate of utterance and duration of excerpt. Lang. and Speech, 6, 1963, 151-164.

Pollack, I. and Pickett, J. M., The intelligibility of excerpts from conversation. Lang. and Speech, 6, 1963, 165-171.

Pollack, I. and Pickett, J. M., Intelligibility of excerpts from fluent speech: auditory versus structural context. Jour. Verbal Learn. and Verbal Behavior, 3, 1964, 79-84.

Silverman-Dresner, T. and Guilfoyle, G., Vocabulary Norms for Deaf Children. Washington, D.C.: The Alexander Graham Bell Association for the Deaf, 1972.

Smith, C. R., Residual hearing and speech production of deaf children. Ph.D. Thesis, City University of New York, 1972.

Stuckey, C.W., Investigation of the precision of an articulation testing program. Jour. Accoust. Soc. of Amer., 35, 1963, 1782-1787.

Subtelny, J. D., The assessment of aural and oral skills: program planning for NTID students. Paper presented at the 1975 meeting of the American Instructors of the Deaf, Greensboro, North Carolina, 1975.

Subtelny, J. D., Assessment of speech with implications for training. In, Childhood Deafness, F. Bess (Ed.), New York: Grune and Stratton, 1977, 183-194.

Thomas, W., Intelligibility of the speech of deaf children. Proc. of International Congress on the Education of the Deaf, Document No. 106, 1963.

Winer, B. J., Statistical Principles in Experimental Design. New York: McGraw Hill, 1972.