THE INTELLIGIBILITY OF DEAF SPEECH TO EXPERIENCED AND INEXPERIENCED LISTENERS

NANCY S. MCGARR
The City University of New York and Haskins Laboratories, New Haven, CT

The present study examined differences between experienced and inexperienced listeners in understanding the speech of the deaf. Listeners heard test words in three conditions: isolated words, followed by isolated and segmented test words. These were contrasts made up of test words which were followed by isolated and segmented test words. However, the test words in each condition were either from the low intelligibility of the test words and the amount of context in the sentence.

Those who work with the deaf are not surprised when a child whose speech is judged relatively intelligible in the classroom is still virtually unintelligible to the "man on the street." That there are judgment differences between experienced listeners (e.g., teachers of the deaf) and inexperienced listeners is widely accepted. In fact, the intelligibility of deaf speech has been rated according to how likely the speaker is to be understood by "most trained teachers of the deaf, most people familiar with deaf speech, or almost everyone" (Thomas, 1963). Several investigators (Brannon, 1964; Markides, 1970; Smith, 1975) have noted that a deaf listener may understand about one word in every five produced by a deaf speaker, while an experienced listener's ability to understand deaf speech seems clearly superior (Mangan, 1961; Markides, 1970; Monsen, 1978; Thomas, 1963). On the whole, intelligibility scores decreased systematically from experienced to naive listeners (Mangan, 1961; Monsen, 1978; Nickerson, 1973; Thomas, 1963). While some overlap in individual has been observed, group scores for naive listeners never approached those of the experienced. Differences between groups have been noted for isolated words as well as for sentences. (Hudgins, 1949; Subtelny, 1977; Thomas, 1963).

The studies cited above were designed primarily to assess speech production in the hearing impaired. However, it is possible to deduce several hypotheses about listener differences from this literature. One hypothesis was suggested by Hudgins and Numbers (1942), who investigated patterns of segmental production in the speech of hearing-impaired children. They suggested that the experienced listener recodes deaf speech to compensate for production characteristics typical of deaf speakers. Since the segmental patterns of deaf speakers are generally unknown to the naive listener, he or she cannot use articulatory cues as an aid in interpretation. However, in their study, and in others that followed (e.g., Smith, 1975), Hudgins and Numbers made no attempt to control for the relative intelligibility of the test items. If their hypothesis is correct, one would expect that the experienced listener would obtain higher scores than inexperienced listeners for test words with relatively low intelligibility compared to those with relatively high intelligibility.

A second hypothesis, also advanced by Hudgins and Numbers, was that experienced listeners made better use of contextual information. They argued that the naive listener was so distracted by the quality of deaf speech that information could not be derived from available contextual cues. On the other hand, higher scores for sentences than for isolated words led Brannon (1964) to conclude that context was extremely important for the naive listener. Thomas (1963) noted that both groups profited from context since scores for "everyday" sentences were higher than those for isolated words.

One difficulty in this area of research is that the term "context" is used in several ways throughout the literature. In the above investigations and in others (Hudgins, 1949; Subtelny, 1977), context was defined as a word produced and heard in a sentence. However, context effects might also include the amount of linguistic information in the sentence, although this factor has not been examined systematically. In Brannon (1964), the context effect also included sentences of different length—thus, the longer the sentence, the more contextual information available to the listener. In addition, the position of the key word scored in a sentence (e.g., in the CID Sentences of Everyday Speech) also may be considered a part of the context effect. For these stimuli, it may be hypothesized that the later a test word occurs in a sentence, the more information the listener has with which to form a correct judgment. However, in the above studies, results emphasizing the importance of contextual effects on listeners' perceptions are complicated by...
contextual effects in the speaker's production of the stimuli. For example, words produced by nondeaf speakers in sentences have been shown to differ in duration, intensity, and so forth from the same words produced in isolation (Lieberman, 1963; McGarr, 1981; Miller, Heise, & Lichten, 1951; O'Neil, 1957; Pollack & Pickett, 1963, 1964). This difference in production has not been examined in the speech of the hearing-impaired. More importantly, in the deaf literature, different vocabulary was used for sentence and for isolated word conditions so that no such comparison is possible.

Finally, in all the studies cited above, the criterion of listener experience was not always carefully controlled. In some instances experienced listeners were very familiar with the children, the speech training protocol, or the test material. In other studies, the listeners were not familiar with any of these factors. The extent to which these variables may account for reported differences between listeners has not been determined.

The purpose of this study thus was to examine two related questions concerning the differences between experienced and inexperienced listeners. First, is the well-known superiority of experienced listeners in decoding deaf speech influenced by the relative intelligibility of test words produced by a deaf speaker? Second, does context benefit the experienced listener more than the inexperienced listener? To examine these questions, the intelligibility of test words was compared in three different listening conditions: sentences, isolated words, and segmented words (originally produced in sentences, excised, and presented to the listeners in isolation).

**METHOD**

*Listeners*

One hundred and twenty listeners participated in the study (60 experienced and 60 inexperienced.) An experienced listener was defined as a person who had more than 1 year's experience in listening to the speech of the deaf. The range was from just over 1 year to 25 years and the mean was 6.8 years. The 60 experienced listeners were teachers of the deaf, speech-language pathologists, and audiologists in schools for the deaf. The listeners did not know the child whose speech they heard; the school at which the child received training, or any aspect of the test material. Each of the listeners had normal hearing and was a native speaker of English.

An inexperienced listener was defined as an individual with no previous experience in hearing the speech of the deaf. The 60 inexperienced listeners were recruited primarily from undergraduate classes. These listeners met all other criteria required of the experienced group.

*Deaf Speakers*

Twenty congenitally and profoundly deaf children from the Lexington School for the Deaf served as sub-

jects. All subjects had been enrolled in the school program for a minimum of 2 years preceding the study. The children were divided equally into two age groups, one of 8-10-year-olds and another of 13-15-year-olds, with five girls and five boys in each group. Subjects had no handicapping conditions other than deafness. The group mean pure-tone average for .5, 1 and 2 kHz was 98.6 dB (ISO) in the better ear. No formal speech evaluation of the children was conducted by the experimenter as part of the criteria for inclusion in the study. However, the speech records of the children were examined by the experimenter, and those children judged by their speech supervisors to have at least average speech intelligibility were selected as candidates. No child whose speech was judged completely unintelligible was considered for the study.

*Materials*

The test materials were 36 monosyllabic words. Each word was empirically defined with respect to its predicted intelligibility when produced by a deaf child. This measure was obtained by ranking all words produced by deaf children in Smith's (1975) study. The 18 monosyllabic words ranked highest for intelligibility and the 18 monosyllabic words ranked lowest for intelligibility formed the test corpus. Scores for these words in the present study were compared subsequently with those of Smith and showed the same clustering of high and low intelligibility scores.

To examine the effect of context, each of the 36 words was embedded in a sentence that varied with respect to the amount of overall contextual information it provided. A measure of high or low contextual information was made for each sentence using a standard word prediction technique. Twenty undergraduates (not listeners) were asked to "fill-in the blank" when presented with a written version of the sentence with the test word omitted. A sentence was defined as high in contextual information if 15 or more undergraduates completed it with the test word. A sentence was defined as low in contextual information if 15 or more undergraduates selected different words to complete the sentence.

The sentences also were designed to vary with respect to other factors believed to be important to listeners: (a) the number of syllables in the sentence, and (b) the location of the test word in the sentence. The sentences were either 3, 5, or 7 syllables in length; the test word occurred either (a) at or near the beginning of the sentence, (b) in the middle of the sentence, or (c) near or at the end of the sentence. For the 36 test words in sentences, all factors (predicted word intelligibility, sentence context, number of syllables, and word position) are relevant to the test material. For the test words in isolation, only predicted word intelligibility is a factor. The test materials are presented in the Appendix.

*Listening Conditions*

Since an isolated word differs from one in a sentence
in both the way it is perceived and the way it is produced, an additional set of stimuli was produced maintaining the same balance of context and word intelligibility. These items originally were produced in sentences but subsequently were heard by the listeners in isolation. These stimuli are referred to as segmented test words and were obtained by processing the audiotape recordings of the children's sentences on the Haskins Laboratories spectrum and waveform editing system. Segmentation was accomplished using both auditory and visual cues.

During input to the computer system, the overall intensity level for each sentence frequently was changed in order to make full use of the available dynamic range of digital storage. Thus, the intensity level of the segmented word frequently was quite different from that of the same word in the sentence. Similarly, productions of a word in isolation also had different intensity levels from productions of the same word in segmented or sentence conditions. Since these differences in intensity alone may account for differences in intelligibility, level recordings of the test word in each condition were made using the Brul & Kjær Level Recorder. The intensity level (in decibels) of each test word in the sentence condition was set as the reference, and the levels for the test word in isolated and segmented conditions were adjusted to this level and rerecorded.

The speech recordings of the children then were prepared to present the following stimuli to the groups of listeners.

1. Test words produced in sentences and presented to the listener in sentences. (Listeners were asked to write down the whole sentence; however, the scores for test words were of primary interest.)
2. Test words produced in isolation and presented to the listener in isolation.
3. Test words produced in sentences, excised from the sentences, and presented to the listeners in isolation (segmented test words.)

In each condition, the deaf speakers' samples were randomized in order to avoid learning effects. That is, each listener heard only one child with no repetition of the same test word on a tape (12 words from each of the 3 conditions described above for a total of 36 items per child).

RESULTS

Six scores were obtained for each child (the 3 stimulus types described above times 2 levels of listener experience). Analyses of variance were performed on these scores to test for significant interactions between listener experience and other factors. Separate analyses were performed for test words in sentences, in isolation, and in segmented conditions. The factors considered in these analyses included: listener experience, predicted word intelligibility, amount of sentence context, and two additional factors pertaining to the speakers—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 performing the analyses of variance, data were transformed using the arcsine transformation (Brownlee, 1965) in order to stabilize the variance. Because of the large number of F tests performed in each of these analyses, the data were interpreted conservatively. Only those effects with a significance level of .01 or less were considered to be statistically significant. Those effects with a significance level between .01 and .02 were considered to be borderline. There was a significant difference between younger and older children for test words in sentences, $F(1, 16) = 11.58$, in isolation, $F(1, 16) = 11.84$, and a borderline significant difference for segmented words, $F(1, 16) = 4.46$. There was no significant difference between male and female subjects for any of the test stimuli.

Listeners' Scores

Table 1 summarizes the mean scores obtained by experienced and inexperienced listeners for each type of speech stimulus. Experienced listeners consistently obtained higher scores than inexperienced listeners. For both groups, scores for test words in sentences were highest, followed by scores for isolated words and then scores for segmented words. Scores for test words in sentences were more than double the scores for segmented words. The greatest difference between listeners occurred on sentences—11 percentage points. This difference was statistically significant $F(1, 112) = 20.5$. In contrast, listeners' scores differed by 6 and 3% for words in isolation and for segmented test words, respectively. While the difference between listeners was statistically significant for isolated words, $F(1, 48) = 14.6$, it was only of borderline statistical significance for segmented words $F(1, 112) = 5.03$. Intelligibility scores also were obtained for all the words in the sentences (cf. Table 1). Scores

<table>
<thead>
<tr>
<th>Type of stimulus</th>
<th>Listeners</th>
<th>Mean score % correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test words produced and heard in sentences</td>
<td>Experienced</td>
<td>.41</td>
</tr>
<tr>
<td>Test words produced and heard in isolation</td>
<td>Experienced</td>
<td>.29</td>
</tr>
<tr>
<td>Test words produced in sentences and heard in isolation (i.e., segmented)</td>
<td>Experienced</td>
<td>.16</td>
</tr>
<tr>
<td>All words produced and heard in sentences</td>
<td>Experienced</td>
<td>.49</td>
</tr>
</tbody>
</table>


based on all words were only slightly higher than scores based on test words alone. There was also a statistically significant difference between each type of stimulus—test words in sentences, in isolation, and in segmented conditions $F(2, 120) = 89.34$.

**Predicted Intelligibility of Test Words**

The mean scores obtained by experienced and inexperienced listeners as a function of the predicted intelligibility of the test words are plotted in Figure 1. Experienced listeners obtained higher scores than inexperienced listeners for both high- and low-intelligibility words in sentence, in isolated, or in segmented conditions. Because the overall pattern of scores was similar for both groups, there was no statistically significant interaction between listener experience and predicted word intelligibility. There was a statistically significant difference between words with predicted high intelligibility and those with predicted low intelligibility. This effect was most pronounced for test words in sentences $F(1, 112) = 22.94$ and in isolation $F(1, 48) = 70$. In these conditions, scores obtained by both groups of listeners were noticeably higher for test words with high predicted intelligibility than those with low. The effect on segmented test words was less, although still statistically significant $F(1, 112) = 15.24$. There was no statistically significant interaction between predicted intelligibility and stimulus type.

**Sentence Context**

The mean scores obtained by experienced and inexperienced listeners for test words as a function of sentence context are plotted in Figure 2. For all conditions, experienced listeners again scored higher on the average than inexperienced listeners, but no significant interaction was found between listener experience and context. The difference between experienced and inexperienced listeners for test words in either high- or low-context sentences was roughly 10%. Since segmented test words originally were produced in sentences, the effect of context on the intelligibility of these stimuli also was examined. The difference between listeners for segmented words produced in high- or low-context sentences was roughly 5%. The magnitude of the context effect is also
evident in Figure 2. Scores for both groups of listeners were greater for the high-context conditions than for the low. Scores for test words in high-context sentences were approximately 16% greater than for those in low-context sentences. This difference was statistically significant $F(1, 112) = 53.61$. For segmented test words, the difference between high- and low-context conditions was approximately 8% for both groups. Thus, the effect of context for words produced and heard in sentences is substantial. If, however, the same test words are segmented, so that although they were produced in context they are heard in isolation, the effect of context is much smaller, and statistically significant $F(1, 112) = 30$.

**Interaction Among Experience, Context, and Intelligibility**

Of special interest was the significant interaction between intelligibility and context for sentences and any interaction involving experience and these factors. The interaction between context and predicted intelligibility ($I \times C$) was statistically significant for test words in sentences $F(1, 112) = 18.5$. For segmented test words, a borderline interaction was obtained for listener experience, context, and predicted word intelligibility ($E \times I \times C$), $F(1, 112) = 5.89$. These three factors are plotted in Figure 3. For test words in sentences, the pattern for experienced and inexperienced listeners was similar, with the difference between groups of listeners averaging about 10% across each of the four combinations of intelligibility and context. For both groups of listeners, the ranking of scores (from highest to lowest) as a function of predicted intelligibility and sentence context were (a) high intelligibility, high context; (b) low intelligibility, high context; (c) high intelligibility, low context; and (d) low intelligibility, low context.

For segmented test words, the overall pattern for experienced and inexperienced listeners show relatively the same effect of intelligibility as was seen in the sentence condition. That is, for both experienced and inexperienced listeners, high-context words were most intelligible and low-context words least intelligible. Also, on the average, scores for test words with high intelligibility were higher than those for test words with low intelligibility. In only one instance did inexperienced listeners receive slightly higher scores than experienced listeners did. For segmented test words with low context, the experienced listeners showed a significant drop in their scores for low-intelligibility words compared to scores for high intelligibility words. This resulted in the borderline interaction cited above.

**Position of the Test Word and Number of Syllables**

An additional analysis of variance was performed to investigate the effects of the position of the test word in the sentence and the number of syllables in the sentence, and to ascertain whether there were any interactions between listener experience and these two factors.

The main effect for position of the test word in the sentence was highly significant $F(2, 35) = 9.26$. No statistically significant effect was found for the number of syllables in the sentence. However, there was a statistically significant interaction $F(4, 35) = 9.16$ between the number of syllables in the sentence and the position of the word in the sentence. Again, there was no statistically significant interaction between listener experience and these factors. Figure 4 plots the significant interaction between the number of syllables and the word position in the sentence for both groups of listeners and also provides insight into the significance of word position. Test words at the beginning of three syllable sentences were less intelligible than those near 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 near (within two syllables) the beginning of the sentence but not in the word initial position. Differences between experienced and inexperienced listeners were greatest for test words near the beginning of five syllable sentences, and next for test words near the middle and finally near the end of seven-syllable sentences. No effects were found for segmented test words.
These data do not support several hypotheses that have attempted to explain the differences between listeners. Hudgins and Numbers (1942) proposed that experienced listeners obtain higher scores than inexperienced listeners because they are familiar with typical errors in production of deaf speech and recode the speech to compensate for these errors. If this were the case, one would expect an interaction between listener experience and predicted word intelligibility. By definition, words with high intelligibility are ones that deaf children are likely to produce correctly. Similarly, words with low intelligibility are ones that deaf children are likely to misarticulate. Hence, if the above hypothesis was correct, experienced listeners would show a greater relative gain for low intelligibility words, since these words should have more errors for the listener to recode. However, no significant interaction was obtained. The measured difference in scores between experienced and inexperienced listeners for test words with high intelligibility was about the same as the difference for test words with low intelligibility, as shown in Figure 1. The lack of a statistically significant interaction between listener experience and predicted word intelligibility does not necessarily mean that experienced listeners recode deaf speech in the same way as inexperienced listeners, but rather that recoding strategies are more subtle and less easily defined than previously proposed.

A second hypothesis (Hudgins & Numbers, 1942; Thomas, 1963) is that experienced listeners simply make better use of contextual cues. Scores for both classes of listeners were higher for sentences with high context than for those with low context (cf. Figure 2), and there was no statistically significant interaction between listener experience and 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 disprove the importance of context, but rather indicates that if there is an interaction effect, it is smaller than has been suggested.

While the effect of context on speech intelligibility has long been realized, Hudgins and Numbers argued that context may be even more important for deaf speech than for "normal" speech. Specifically, they hypothesized that the effect of articulatory errors on the intelligibility of deaf speech could be reduced by the contextual information of the sentences, and conceivably, where the articulatory errors are greatest, context becomes more important. This hypothesis of an interaction between intelligibility and context was supported by the data. The effect of word intelligibility accounted for a greater change in scores for high-context sentences than for low-context sentences. While there was a significant interaction between 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. This effect was observed even for individual children whose intelligibility scores were low (< 30%) (cf. McGarr, 1978). These results contradict Sitler,

---

**DISCUSSION**

Scores for the experienced listeners were consistently higher than those for inexperienced listeners. Further, the differences in the test scores between experienced and inexperienced listeners were essentially equal across all factors investigated: (a) Predicted word intelligibility, (b) degree of sentence context, (c) number of syllables in the sentence, and (d) position of the test word in the sentence. For both groups of listeners, the scores for test words in sentences were consistently higher than scores for test words in isolation. The lowest scores were those for segmented words.
Schiavetti, and Metz (1983), who found no effect of context for subjects with poor intelligibility. It should be noted that Stiller et al. did not control for the degree of context in their test materials and also used different vocabulary for their isolated words and sentences.

Another view 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 or her scores would be lower. In the literature, definitions of experienced listener included persons who knew the subjects, such as teachers or parents (Mangan, 1961), listeners who were trained on either the test materials or the deaf speakers (Hudgins, 1949), as well as 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. Although the study did not examine differences between experienced listeners who knew the subjects and those who did not, the hypothesis that personal knowledge of the speaker alone enables the experienced listener to obtain higher intelligibility scores was not supported (see also Gulian & Hinds, 1981). While it is likely that children who are known to parents or teachers may be more intelligible to them than to other listeners, further research is warranted to quantify the effect of personal knowledge.

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 an 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 factor investigated implies that the effect of experience is not due to any superficial recoding of deaf speech on the part of the listener. If the factors considered in this study (i.e., context, predicted word 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.

In addition to the main effects tested, it is also known that the difference between experienced and inexperienced listeners was not due to any secondary effects such as idiosyncrasies in particular children or in specific test words. Overall scores for younger children were slightly poorer than those for older children, as was also observed by Smith (1975), and there was little difference 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.

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 study suggests that the advantage of experience cannot be attributed simply to one or two variables, at least for the factors considered within this study. Consequently, the differences between experienced and inexperienced listeners must be due to fairly complex aspects of deaf speech that are not immediately apparent to the listener. The fact that the difference between listeners was constant suggests that the effect occurs fairly 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.

ACKNOWLEDGMENTS

This research was conducted in partial fulfillment of the requirements for the Ph.D. at the City University of New York. The author wishes to acknowledge the very helpful comments of Harry Levitt, Katherine S. Harris, and Vicki Hanson. Preparation of the manuscript was supported by Grants NS-13617 and 13870 to Haskins Laboratories, and Grant NS-17764 to CUNY.

REFERENCES


NICKERSON, R. S. Computerized speech-training aids for the
APPENDIX

Test sentences recorded by the deaf subjects. The test word is italicized in each sentence. These words were also produced in isolation.

HIGH CONTEXT

3 Syllables

Keep quiet.
Read the book.
Come with me.
The dog barks.
That’s no good.

5 Syllables

The cat chased the mouse.
My name is Nancy.
Get your coat and hat.
Get your ball and bat.
Did you brush your teeth?
Is there no more milk?

7 Syllables

That man is not my father.
I wish I had a pony.
We have food for the picnic.
The flag is red, white and blue.
May I have a piece of cake?
Can you dive in deep water?

LOW CONTEXT

3 Syllables

Feed the dog.
Have a lot.
You did it.
I need it.
Get the cake.
This is his.

5 Syllables

They will come again.
Is that the tall one?
Mother has the car.
Who wants this ice cream?
It’s easy to hear her.*
He said he could go.

7 Syllables

The book is on the table.
What was the name of that boy?
If it’s cool I cannot go.
Is the fat baby crying?
It is nice on a fall day.
We will go to the beach today.*

*These sentences contain an additional syllable.

Reprinted from Journal of Speech and Hearing Research
September 1983, Vol. 28, No. 3
Copyright © 1983 by the American Speech-Language-Hearing Association

INFORMATION ON REPRINTS AND PERMISSIONS

The appearance of the fee codes in this journal indicates the copyright owner’s consent that copies of articles may be made for personal or internal use or for personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per article fee of $1.00 through the Copyright Clearance Center, Inc., 21 Congress Street, Salem, Massachusetts 01970, for copying more than one copy as indicated by Sections 107 and 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as copying for general distribution, advertising or promotional purposes, for creating new collective works, or for resale. In these cases, requests for permission to reprint and/or quote from the journals of the Association must be obtained from the American Speech-Language-Hearing Association and from the individual author or authors of the material in question.

Consent is extended for copying articles for classroom purposes without permission or fee unless otherwise stated in the article.