A lip-reading assessment for profoundly deaf patients

by

L. F. A. Martin, G. M. Clark, P. M. Seligman and Y. C. Tong
(Melbourne, Australia)

Introduction

To understand spoken sentences, first the acoustic information is processed, and secondly linguistic knowledge is applied (Fry, 1961; Kalikow et al., 1977). The more the spoken message contains linguistic redundancies in the form of lexical, syntactical and semantic constraints, the less the listener needs to rely on processing the details of the acoustic signal. For normal listening conditions there is usually enough information available to make an unambiguous decision about the spoken message. However, when the acoustic signal is degraded, more reliance is placed on the context in which the message was spoken (Miller et al., 1951). For some hearing-impaired individuals the auditory signal is permanently degraded. In addition, if people are totally or profoundly deaf, speech is usually perceived via lip-reading alone. This is usually difficult because not all phonemes can be clearly distinguished visually. For example, some phonemes form homophenous groups, i.e. they look the same on the lips; such a group would be the bilabial plosives and nasal /p, b, m/. Since the information reaching the individual is incomplete, greater reliance must also be placed on linguistic skills and on the context in which the message is spoken.

Attempts have been made to quantify linguistic skills. For example, Duffey and Giolas (1974) have measured the redundancy of traditionally used sentences, and Kalikow et al. (1977) constructed sentence lists in which linguistic context is a variable under test (Speech Intelligibility in Noise sentence test or SPIN test). With the SPIN test the subject is presented with sentences in which the last word can be predicted either easily or with difficulty from the preceding words. For example, an easily predictable last word would be ‘map’ in the sentence ‘We’re lost so let’s look at the map’. Here ‘lost’ and ‘look’ are semantically linked to the word ‘map’, while ‘at the’ points syntactically to the last word as being a noun. On the other hand, a poorly predictable last word would be ‘map’ in the sentence ‘I should have considered the map’. Each sentence list consists of 25 high predictability and 25 low predictability sentences, mixed together in irregular order, and the patient is asked to decide the last word in each sentence.

The sentences were designed so that they were of approximately equal length and relatively short, so that a subject’s memory limitation would not confound the responses. The phonetic composition of the final words and the content words of the high predictability sentences were representative of everyday American English and were not weighted in favour of visually distinctive phonemes so that the subject’s lip-reading abilities could be assessed.
The present study was undertaken to assess high and low predictability word discrimination in profoundly or totally deaf patients under different conditions using the SPIN sentences.

The different conditions were lip-reading alone, and lip-reading in combination with a prosthetic device. For one subject the prosthesis was a multiple-channel cochlear implant; for four subjects it was a powerful hearing aid, and for two other subjects a simple vibrotactile device was used.

**Methods**

**Subjects**

Eight profoundly deaf adults with post-lingually acquired hearing losses participated in this study. The salient features of each patient’s case history are presented in Table I. For those patients who had a sudden hearing loss the aetiology was known. This was meningitis for P3, P6 and P8, and a car accident for P1. For the other patients the cause of deafness could not be determined with any certainty. The average age was 52 years.

All patients had profound bilateral sensorineural hearing losses. Five patients (P1, P3, P5, P6 and P7) had no responses to pure tone air conduction testing at the limits of the audiometer, for the frequencies 125–8,000 Hz. The results of the other three patients (P2, P4 and P8) are summarized in Table II.

**TABLE I**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sex</th>
<th>Age</th>
<th>Onset</th>
<th>Years of deafness</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>M</td>
<td>48</td>
<td>Sudden</td>
<td>4</td>
</tr>
<tr>
<td>P2</td>
<td>M</td>
<td>63</td>
<td>Gradual/sudden</td>
<td>15</td>
</tr>
<tr>
<td>P3</td>
<td>F</td>
<td>55</td>
<td>Sudden</td>
<td>38</td>
</tr>
<tr>
<td>P4</td>
<td>M</td>
<td>45</td>
<td>Gradual</td>
<td>16</td>
</tr>
<tr>
<td>P5</td>
<td>F</td>
<td>63</td>
<td>Gradual</td>
<td>12</td>
</tr>
<tr>
<td>P6</td>
<td>F</td>
<td>30</td>
<td>Sudden</td>
<td>19</td>
</tr>
<tr>
<td>P7</td>
<td>F</td>
<td>60</td>
<td>Gradual</td>
<td>30</td>
</tr>
<tr>
<td>P8</td>
<td>M</td>
<td>54</td>
<td>Sudden</td>
<td>40</td>
</tr>
</tbody>
</table>

Average age 52.

**TABLE II**

<table>
<thead>
<tr>
<th>Hz</th>
<th>P2 R</th>
<th>P2 L</th>
<th>P4 R</th>
<th>P4 L</th>
<th>P8 R</th>
<th>P8 L</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>90</td>
<td>75</td>
<td>85</td>
<td>90</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>115</td>
<td>95</td>
<td>105</td>
<td>115</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,000</td>
<td></td>
<td>105</td>
<td>115</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

— Indicates no response to the maximum output of the audiometer.

Only P4 was a regular hearing aid user; he wore a powerful body aid in the right ear. The patient P1 had had a multiple-channel cochlear implant inserted into his left ear. This device has been described in detail (Tong et al, 1979, 1980). Briefly the implant electrically stimulates the auditory nerve via an array of electrodes that is placed in the scala tympani. Information and power are transmitted to the implant from a speech processor which processes the acoustic signal. Speech processing is aimed at providing information about the second formant and fundamental frequency. Another patient, P2, also had a multiple-channel cochlear implant operation on his left ear, but at the time of testing this device had a malfunction and was not used.

**Stimuli**

Two lists (List 2.1 and List 2.8) of the SPIN test sentences (Kalikow et al., 1977) were videotaped. The speaker was a female audiologist with an English accent. An audio and visual recording was made in an anechoic chamber on a reel to reel IVC video taperecorder, using a Sony DXC 1600P
colour video camera. The image of the speaker’s face on the monitor screen was adjusted to be approximately life size. The microphone used was a Superscope cardioid condenser EC-7.

Results

Experiment 1

In this study List 2.1 was presented to all eight patients as a lip-reading test. Each patient received the test individually and sat approximately four feet from the monitor screen. They were asked to repeat the last word of each sentence; this was recorded by an audiologist who sat with the patient during the test. In scoring the correct words credit was given, if the patient’s response represented the stem of the word but a morphological appendage was omitted or added. For example sentence 15 is: ‘Crocodiles live in muddy swamps’; if the patient replied with ‘swamp’ this was counted as correct.

Table III shows the results for this lip-reading test. There are 50 sentences in the list so that each patient scored from a possible total of 50 words. Twenty-five words were from high-predictability sentences and 25 from low-predictability sentences.

The average total scored was 13.1 words. The average high-predictability score was 10.3 and the average low predictability score was 2.9 words. Applying the ‘t’ test for paired observations to the predictability scores showed there was a significant difference at the 0.01 level of significance, in favour of the high predictability scores being larger than the low predictability scores (t = 5.02).

Experiment 2

In this study P1 was tested with List 2.8. He was allowed to use the multiple-channel cochlear implant and speech processor to receive the auditory signal as well as to lip-read. This test was administered approximately 15 months after the patient P1 had started to use a speech processor. List 2.1 and List 2.8 were presented on different days but within one week of each other. Both lists were again presented under the same conditions approximately six months later to ascertain whether any improvement could be measured using the speech processor. During these six months the patient had the first opportunity to take a wearable speech processor home, and to use it in his normal day-to-day life.

The results of this study are presented in Table IV. From these results it can be seen that P1 scored 18 per cent and 16 per cent for lip-reading alone, and 60 per cent and 54 per cent for lip-reading and the speech processor.
The point of interest from this study is whether the 'lip-reading alone' scores can be considered as significantly different from the 'lip-reading and speech processor' scores. Since the sample size is equal to one, conventional statistical analysis is difficult to apply. However, Thornton and Raffin (1978) have produced a statistical model for comparing scores for two different test lists with the same subject. From their model for a first score of 18 per cent, a second score exceeding 34 per cent can be considered to be significantly different from the first score at the 95 per cent level of confidence. Using the data provided by Raffin and Thornton (1980), it can be seen that the difference between lip-reading alone, and lip-reading plus the speech processor is significant at the 0.0001 level of confidence, i.e. such a difference will occur by chance one time out of 10,000. This is true for both presentations of the two lists.

When looking at the two 'lip-reading alone' scores (18 per cent, 16 per cent) these are different at the 0.7872 level of confidence, i.e. chance occurrence is 78 times out of 100. Therefore these can be considered to be not significantly different; the lip-reading performance neither improved nor deteriorated over the six-month interval. Similarly the level of confidence for a significant difference between the two scores for lip-reading combined with the speech processor (60 per cent, 54 per cent) is 0.5419. Therefore the performance for lip-reading and the speech processor did not change in six months.

Experiment 3

In this study List 2.8 was presented to P4, P7 and P8 when they were wearing a hearing aid as well as lip-reading. The patient P4 was a regular hearing aid user, but testing had failed to reveal any speech discrimination using his residual hearing alone. He scored 0 per cent for open-set AB words (Boothroyd, 1968) at 120 db. SPL under headphones. During a hearing aid evaluation this patient did not score significantly better than chance for tests using a closed set of five spondees, or a closed set of five sentences. He did perform fairly well on a test using a closed set of 10 environmental sounds, scoring 7 out of 10. It was therefore interesting to find out whether the hearing aid contributed to lip-reading performance or whether the main benefit was monitoring environmental sound and forming an auditory link with the world. On the other hand, P7 and P8 were given the test as part of a hearing aid evaluation; it was the first time for many years they had experienced amplified sound. Again no speech discrimination score was obtained for AB words under headphones (0 per cent for P7 at 125 db. SPL and 0 per cent for P8 at 130 db. SPL). Neither patient obtained a score significantly different from chance (at the 0.05 level of confidence) for a closed set of five spondees when using a hearing aid. Also P7 had a similar non-significant score for a closed set of five sentences when using a hearing aid. Thus it was of interest to see whether any immediate benefit would come about when information from the hearing aid was combined with lip-reading.

Both test lists 2.1 and 2.8 were presented in an anechoic chamber to reduce the possibility of ambient noise masking out the speech signal in the lip-reading and hearing aid condition. Table V shows the aided thresholds for these patients.

The results for these three patients are shown in Table VI. Applying the statistical

<table>
<thead>
<tr>
<th>Hz</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>750</th>
<th>1,000</th>
<th>1,500</th>
<th>2,000</th>
<th>3,000</th>
<th>4,000</th>
<th>6,000</th>
<th>8,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td>81</td>
<td>83</td>
<td>68</td>
<td>61</td>
<td>67</td>
<td>66</td>
<td>64</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td></td>
<td>72</td>
<td>58</td>
<td>51</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td></td>
<td></td>
<td>85</td>
<td></td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hz</th>
<th>900</th>
<th>1,200</th>
<th>1,600</th>
<th>2,200</th>
<th>3,200</th>
<th>4,000</th>
<th>6,000</th>
<th>8,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td></td>
<td></td>
<td>82</td>
<td>78</td>
<td>85</td>
<td>73</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>P7</td>
<td></td>
<td></td>
<td>72</td>
<td>78</td>
<td>85</td>
<td>73</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>P8</td>
<td></td>
<td></td>
<td>73</td>
<td>84</td>
<td>85</td>
<td>73</td>
<td>84</td>
<td>84</td>
</tr>
</tbody>
</table>
A LIP-READING ASSESSMENT FOR PROFOUNDLY DEAF PATIENTS

TABLE VI
THE NUMBER OF WORDS CORRECT, SCORED BY THREE PATIENTS FOR TWO SPIN TEST SENTENCE LISTS UNDER CONDITIONS OF LIP-READING ALONE AND LIP-READING COMBINED WITH A HEARING AID

<table>
<thead>
<tr>
<th>Lip-reading alone</th>
<th>Lip-reading and hearing aid</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td>7 (6H 1L)*</td>
</tr>
<tr>
<td>P7</td>
<td>11 (8H 3L)</td>
</tr>
<tr>
<td>P8</td>
<td>19 (14H 5L)</td>
</tr>
</tbody>
</table>

* The numbers in brackets indicate components of the total score attributable to high predictability sentences (H) and to low predictability sentences (L).

model of Raffin and Thornton (1980) for P4, his scores are significantly different at a confidence level of 0.0037.

However, for P7 and P8 the score for the two conditions cannot be considered significantly different at the 0.05 or the 0.01 level of confidence.

Experiment 4

This study was similar to Experiments 2 and 3 inasmuch as List 2.8 was presented in an aided lip-reading condition. In this instance the aid was a simple vibrotactile device consisting of a powerful body aid attached to a bone vibrator. When held in the fingertips the bone vibrator provides vibrations in response to sounds. This would provide some information about prosody, as well as (mainly) duration and intensity of the speech signal.

Two patients (P2 and P3) participated in this study; this test was part of a number of tests administered as an initial evaluation of this device. The patients were also tested in an anechoic chamber since ambient noise is easily amplified to the bone conductor. It was thought that this aid could provide an alternative source of help for lip-reading for profoundly deaf patients who are unable to gain benefit from conventional hearing aids.

The results are shown in Table VII. For patient P2 the level of confidence for the two scores being significantly different is 0.64 (Raffin and Thornton, 1980). Thus they can be said reasonably to have come from the same population; that is, the vibrotactile device did not seem to help lip-reading. Similarly for patient P3 the confidence level for a significant difference between the two scores is 0.85. Again it is reasonable to accept that there is not a significant difference between the two scores.

Discussion

Experiment 1

The results from Experiment 1, where all patients were tested for lip-reading, show that the sentences with contextual information facilitate the intelligibility of the last word more than those that have no context, when patients have to lip-read. This is in keeping with traditional views about lip-reading (Jeffers and Barley, 1981). Lip-reading the last words of the low-predictability sentences is like lip-reading words, but with a variable carrier phrase. Therefore, the perceptual and speech processing mechanisms do not have the benefit of either semantic contextual information or a predictable carrier phrase.

Examination of the errors made by patients shows that they substituted homophenous words and speech sounds for the stimulus words. It was also found that only 9 per cent of errors produced indicate a failure by the patient to segment the word from the rest of the sentence; this occurred when patients responded with a two-syllable word rather than a monosyllabic word, e.g. ‘races’ for ‘crates’.
Another point of interest is that the test appears to be fairly difficult, the range of scores being 6–23 words correct. The difficulty of the test would have to be borne in mind if it were to be administered routinely to patients who were not used to this type of testing and were not aware of the significance of the results.

Experiment 2

In this experiment the scores from lip-reading combined with the speech processor and multiple-channel cochlear implant were compared to lip-reading alone. By use of the Thornton and Raffin model it was shown that the ‘speech processor and lip-reading’ condition was significantly better than lip-reading alone, and thus the cochlear implant has succeeded in improving the patient’s ability to communicate. However, when comparing the scores over a six-month interval, neither of the two conditions changed significantly. Thus it can be concluded that the six months of experience with the wearable speech processor did not improve performance either with lip-reading or with the ability to use the speech processor. This is not surprising when it is considered that the patient had been using a laboratory-based speech processor fifteen months prior to taking the wearable one home. Any improvement in lip-reading or ability to use the speech processor could have taken place within that period.

Examining the scores for high- and low-predictability sentences, it can be seen that the high-predictability sentences contribute more to the intelligibility of the last word for both conditions. However, for lip-reading combined with the speech processor, the low-predictability words also improve, as do the high-predictability words. This improvement can be attributable to the extra auditory information from the speech processor combining with the visual information. It would be interesting to know what contribution is made to intelligibility from prosodic and from segmental information.

Experiment 3

In this study, which assessed hearing aid performance, only P4 showed a significant improvement in the ‘lip-reading and hearing aid’ condition over the lip-reading-alone condition. Possibly prosodic information was provided by the aid since very poor scores were obtained with open-set and closed-set word tests.

The approximate three-fold improvement of ‘lip-reading plus hearing’ over lip-reading alone compares favourably with the results with the cochlear implant (P1). Therefore this test may have a role to play in the pre-operative assessment of cochlear implant patients.

The fact that similar scores for both conditions were obtained for P7 probably indicates that some learning to interpret the amplified sound is necessary before a patient can make full use of the hearing aid. Evidence for this comes from the fact that no speech discrimination score was obtained using the hearing aid alone, and also that the patient described the sound of speech as a noise. In this case the patient would need a long hearing aid trial and another assessment to ascertain whether any improvement could be obtained with the aid before considering a cochlear implant.

The results for P8 show that he also scored slightly worse for the aided lip-reading condition compared with lip-reading alone. This patient was not used to amplification and reported that it disturbed his concentration when lip-reading. This patient would also need an extensive hearing aid trial.

Experiment 4

Risberg (1974) has shown that, with a known set of 15 sentences, a normally hearing subject could improve from 42 per cent correct for lip-reading alone to 74 per cent correct for lip-reading plus a hand-held vibrator. Similarly Plant (1979) has shown that the use of a hand-held vibrator can improve lip-reading scores by approximately 15 per cent. The patients in this study showed no such improvement. This could be due to a lack of experience with the supplementary information, although a subsequent trial with the aids led both subjects to reject them as lip-reading aids.
The results with these two patients and with patients P7 and P8 suggest that experience both with the aid and possibly previous lip-reading experience may be important in benefitting from lip-reading supplements.

Conclusion

The results obtained in this study show that the SPIN test sentences, when presented as a lip-reading test, could be a useful means of evaluating profoundly deaf patients for a multiple-channel cochlear implant operation. Sentence materials with controlled context are a useful means of analysing each patient’s lip-reading skills. Patients who do exceptionally well with the highly predictable sentences would be expected to manage fairly well with everyday conversation. Sentence tests are generally considered to be better indicators of conversational competence than other types of audiological tests. Since lip-reading is the profoundly deaf person’s principle means of communicating, it is important to assess this together with any possible lip-reading supplements, as part of any pre-operative evaluation.

The results are also interesting from the point of view of assessing a profoundly deaf person’s performance with a hearing aid. It has been well-established that many profoundly deaf people can benefit from a hearing aid (Fujikawa and Owens, 1978, 1979), and that in some cases a patient’s performance with a hearing aid exceeds that or is comparable to that with a single-channel cochlear implant (Owens and Telleen, 1981). In the present study P4’s performance with a hearing aid was comparable with the results obtained with P1 using a multiple-channel cochlear implant. For such patients it is not difficult to assess that the likely benefits of a multiple-channel cochlear implant may not exceed their performance with a hearing aid. Similarly, for patients who cannot tolerate amplification, for example those with excessive recruitment, Tullio’s effect or tinnitus problems, it is clear that a hearing aid will not be of benefit.

However, for patients who show no improvement with a hearing aid but who can tolerate amplification, it is difficult to decide what the benefits of a multiple-channel cochlear implant are in comparison to a hearing aid. This may be because, at the time of testing the hearing aid, the patients may not have worn any amplification for many years. This was the case with P7 and P8 in this study and has also been observed by Fujikawa and Owens (1978). Therefore more knowledge is needed about how such patients may learn to interpret information from the hearing aid and whether in time the hearing aid can help to improve speech intelligibility and supplement lip-reading. Of course, the hearing aid may give the patient benefits other than improvements in speech communication, by allowing him to monitor environmental sounds and providing an auditory link with the world. However, it is felt that it is important to establish criteria of speech intelligibility by which to compare a multiple-channel cochlear implant with the patient’s performance with a hearing aid, and to provide the patient with information by which he can make his choice.

Acknowledgements

The Department of Science and Technology of the Australian Government and the National Health and Medical Research Council of Australia provided financial support.

Mr. Greg Cook provided technical support and Mrs. Kerrin Morris and Mrs. Enid Utton provided secretarial assistance. Mr. Richard Dowell assisted in the administration of some of the tests. The National Acoustic Laboratories (Melbourne) and Optical Prescriptions and Spectacle Makers Pty Ltd provided hearing aids for this study.

References


Address for correspondence:
Professor G. M. Clark,
Department of Otolaryngology,
University of Melbourne,
The Royal Victorian Eye and Ear Hospital,
East Melbourne 3002,
Victoria.