Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 502–505
(S. Karger, Basel/New York 1965).

Recoding Speech for the Deaf and Hard of Hearing

By Arne Risberg, Stockholm

For the profoundly deaf and the severe hard of hearing lip-reading is the principal means of comprehending verbal communication. Some individuals acquire very good results in this art but for others, and especially for the congenitally deaf, lip-reading is a very poor supplement for hearing.

During the last years many investigators have tried to find out what specific ability characterizes a good lip-reader. The results are sometimes contradictory. Other investigators have applied the theory and the methods of modern structural linguistics to the lip-reading situation. The aim has been to find the visual cues to the phonemes. *Woodward* and *Barber*[10] studied the visual discriminability of the English consonant phonemes in $C_1V \cdot C_2V$ combinations, and their results show that the consonant phonemes can be grouped in 4 units comprising the following phonemes. Unit 1: p b m; Unit 2: W, w, r; Unit 3: f, v; Unit 4: t d n l θ ð s z č j š ž y k g h. Discrimination is possible between the units but not within them. The 24 different phonemes in English have thus been reduced to only 4 for the lip-reader. The fact that lip-reading is possible indicates that the lip-reader must read on a higher level than the phoneme and also that he must make use of the redundancy in the language in a very efficient way.

Numbers[5] and others have shown that a combined use of residual hearing and lip-reading gives better results than only lip-reading even if hearing alone cannot be used for comprehending speech. The explanation seems to be that the information transmitted by hearing gives data to reduce some of the uncertainty in the visual message.

The information that can be transmitted by a very small residual hearing is very limited and it is probable that only amplifying the

speech sounds is not the most efficient way of transmission. The development of techniques for automatic analysis of speech parameters such as voicing, fricative, pitch, and so on, gives us a new method for transmitting these parameters to the deaf and hard of hearing. The parameters can be recoded to signals that are better perceived than the mere amplified sounds. The parameters can either be recoded into new acoustic signals, or into tactual and visual forms. How this recoding shall be done to get the best possible result for a specific type of loss of hearing must be studied first by investigating the coding dimensions that can be used for the different senses and then by building practical recoding systems. In some cases when the residual hearing is relatively good it is possible to recode speech so that simultaneous lip-reading is not necessary. It is less likely that a recoding of speech to tactual or visual signals can give sufficient information but it is possible that some individuals after long training can perceive everyday speech at a reduced speaking rate and tactual recoding can then be used for communication with the deaf and blind.

The recoding of speech after these principles can either be used as an aid in speech perception or as a means for speech correction. In both instances the instrumentation and the technique is basically the same. In speech correction work the subject's own voice is processed in the same way as that of the therapist's and the perceived visual, tactile, or auditive patterns serve as a feedback for correcting the pronunciation of sounds and words. As a device for speech reception the auditive and tactual channels are of primary interest since the visual channel would interfere with lip-reading. The personal aid should also be a light-weight portable system. No such limitations need to be considered in speech training.

Experiments with recoding of speech for the deaf and hard of hearing of different degrees of loss of hearing have been made at several laboratories[2, 4, 7, 9]. Some of these experiments have been directed towards speech correction only[1]. Very few results have been published.

At the Speech Transmission Laboratory these problems have been taken up for a general study. The work on tactual speech transmission published by *Pickett* at the Fourth International Congress of Phonetic Sciences in Helsinki, 1961, and in later publications[6] is continued and extended. Visual speech transmission using simple instantaneous spectrum analyses for speech correction purposes have

been tried with hard of hearing subjects and normal hearing children with speech defects. The results are very promising.

The technical problem of extracting signals corresponding to the distinctive features[3] is studied and possible coding dimensions for these signals both as tactual, visual, and auditory signals are investigated.

Results obtained were reviewed at the congress.

References

1. *Anderson, F.:* An experimental pitch indicator for training deaf scholars. J. acoust. Soc. Amer. *32:* 1065–1074 (1960).
2. *Gault, R. H.:* The interpretation of speech by tactual and visual impression. Arch. Otolaryng. *4:* 228–239 (1926).
3. *Jakobson, R.; Fant, G.* and *Halle, M.:* Preliminaries to speech analysis. The distinctive features and their correlates. Acoustics Laboratory, M.I.T., Technical Report No. 13 (Jan. 1952).
4. *Johansson, B.:* A new coding amplifier system for the severely hard of hearing; Proc. 3rd int. Congr. Acoustics, Stuttgart 1959; vol. 2, pp. 655–657 (Elsevier, Amsterdam 1961).
5. *Numbers, Mary E.* and *Hudgins, C. V.:* Speech perception in present day education for deaf children. Volta Rev. Reprint No. 599 (Sept. 1948).
6. *Pickett, J. M.:* Transmitting speech sounds by a tactual vocoder and by lip-reading. Royal Institute of Technology, Speech Transmission Laboratory Report No. 27 (March 1963).
7. *Pimonow, L.:* L'application de la parole synthetique dans la correction auditive. Acustica *12:* 285–290 (1962), English translation J. Aud. Res. *3:* 73–82 (1963).
8. *Potter, R. K.; Kopp, G. A.* and *Green, H. C.:* Visible Speech (Van Nostrand, New York 1947).
9. *Wedenberg, E.:* Auditory training of severely hard of hearing using a coding amplifier. Proc. 3rd int. Congr. Acoustics, Stuttgart 1959, vol. 2, pp. 658–660 (Elsevier, Amsterdam 1961).
10. *Woodward, Mary F.* and *Barber, Caroll G.:* Phoneme perception in lipreading. J. Speech and Hearing Res. *3:* 212–222 (1960).

Author's address: Mr. A. Risberg, Speech Transmission Laboratory, Royal Institute of Technology, *Stockholm 70* (Sweden).

Discussion

Delattre (S. Barbara): If a 1000 cps low pass filter still allows a distinction such as [t/s] to be made, it must mean that there exist cues in the first and second formant transitions that really count. I can think of two, offhand. The first transition has a higher locus and a slower tempo, and the second formant has a lower locus, for [s] than for [t].

Denes (Murray Hill): Your last slide, showing the high articulation scores for distinguishing between /s/, /st/ and /t/, for speech low-pass filtered at 700 c/s, surprised me greatly. The voice energy for these sounds is largely concentrated at frequencies well above this cut-off and I wonder if you could say how sharp your low-pass filter was and also what acoustic feature you think the subjects utilize in making these distinctions.

With regards to the question of the value of a non-portable re-coding device for aiding in speech recognition of the deaf, such devices would be useful with telephones. During normal face-to-face conversation the deaf person can supplement his deficient hearing by lip-reading; this is not possible when he uses the telephone. Under the conditions a re-coding device, with a suitable display, could take the place of lip-reading and make speech communication possible even if the display of the re-coding device on its own were not sufficient.

Lebrun (Bruxelles): In answer to the question "How can deaf or hard of hearing people who lip-read distinguish between [p], [b], and [m]?" I should like to suggest that perhaps they use the following cue:

When [p] is pronounced, the lips of the locutor – at least, if he speaks French – are pressed harder against one another than when [b] or [m] is pronounced. On the other hand, when [b] is articulated, the lips are slightly protruded, whereas they are slightly retracted when [m] is uttered. The slight protrusion of the lips during the "tenue" of the [b] (or [p]) may to some extend anticipate the explosion which is to follow, or else result from the relatively high pressure within the mouth. In the case of [m], the slight retraction of the lips may perhaps be due to the relatively low pressure in the mouth cavity.