VISUAL INFORMATION AND SPEECH ACQUISITION OF THE DEAF

DIRK-JAN POVEL

Psychological Department
University of Nijmegen
P.O. Box 9104
6500 HE Nijmegen, The Netherlands

BEN MAASSEN

Interdisc. Inst. of Child Neurology
University of Nijmegen
P.O. Box 9101
6500 HB Nijmegen, The Netherlands

ABSTRACT

Substantial improvement of speech of the deaf can only be obtained if we succeed in providing them with additional information concerning speech through other than auditory channels. Since there have been many attempts at developing visual aids but very little success [2], we consider some basic issues involved in the development of visual aids for speech training and propose a number of assumptions that guide our undertaken to construct an effective visual aid.

INTRODUCTION

Currently we are involved in a project in which we are developing computer controlled visual aids displaying acoustic information of speech to be used in speech acquisition of the deaf. The basic motivation behind the project is the belief that the only way to substantially improve the results of speech training of the deaf is by introducing (visual) aids that supply additional information and feedback about speech, which can be incorporated in a speeech training program. We belief that the deaf child can only form an adequate and stable internal representation of speech if we supply them information that shows what speech looks like and provides them with an opportunity to examine the results of all sorts of articulatory gestures and to check the successfulness of attempts to produce specific speech acts. Because of the enormous growth in recent years of computational power and graphics facilities, it is now possible to design all sorts of visual aids, evaluate them in a training situation and subsequently adjust the design on the basis of this evaluation, in a most flexible way.

However, the more possibilities there are, the more decisions have to be taken about different design aspects of the aids. Therefore, we present in this paper a preliminary framework that allows to see the different dimensions of the problem and indicates the type of questions that should be asked (and answered).

TWO BASIC QUESTIONS

To develop a visual aid means to answer two questions: WHAT information to display and HOW to display it. Attempts to answer these questions lead to the most fundamental aspects of speech perception and production as well as to basic questions concerning the essential differences between the processing performed by the ear and the eye. The answer to these questions is in part also determined by the view one holds with respect to the method of speech education of the deaf. Although educational aspects will undoubtedly play an important role in the ultimate form of the aids, in this paper we will not deal with these aspects but confine ourselves to the more fundamental issues relating to how relevant aspects of speech can be made visible for speech training purposes.

THE ULTIMATE GOAL: THE IDEAL SPEECH VISUALIZER

What would be the ultimate goal of a project like this? As we see it, the ideal speech visualizer for speech training should completely take over the missing auditory function of the deaf child.

In order to attain this ideal two problems should be solved. First, we must find ways to present the acoustic information about speech in a form which is digestible by the eye. In view of the great differences between the way information is processed by the ear and the eye, it will be necessary to do a lot of preprocessing that transforms the acoustic information into a form suitable for the eye. Secondly, we will have to find ways to inform the child how the visual display relates to speech. For the deaf child acquiring speech needs to understand how different visual dimensions and combinations of these dimensions are related to speech production. Even more importantly, information must be supplied about how these dimensions are used to produce different speech-related acts. This means that the display must also present normative information. We will return to this point below.

To what extent this ideal can be realized is at present unpredictable.

But even a superficial study of the differences between visual and auditory perception on the one hand and of the coding of speech in the acoustic signal on the other, makes clear that this is a most complicated undertaken.

A MORE PRACTICAL APPROACH: SOME BASIC ASSUMPTIONS

In order to make the problem somewhat more manageable we will conceive of the speech signal as describable in terms of a limited number of parameters that are related to basic aspects of speech production. The traditionally used parameters associated with intensity, fundamental frequency, timing and spectral composition seem most useful because they can, after some transformations, be related to the basic aspects of speech production: respiration, phonation and articulation. As a first step towards developing a theoretical frame we will formulate some assumptions with respect to the way acoustic information about speech should be mapped in the visual mode. These assumptions form together the starting-point for our approach followed in this project. As such, the assumptions can be seen as requirements for the visual aids to be developed.

Assumption 1

There must be a unique and fixed relation between acoustic and visual parameters. This assumption is based on the consideration that a stable internal representation can only be formed if different dimensions of the process are uniquely connected to specific aspects of the visible output. It implies that one should not use the same diagram to display different acoustic parameters. It also implies that if a certain acoustic feature is once associated with some visual attribute this should not be changed later.

Assumption 2

The visually presented information concerning speech should be as complete as possible. The following arguments form the basis for this assumption. In the first place having all information available concerning a skill to be developed is the natural situation, independent of the question whether the subject uses all the information all the time. Secondly, there is a practical argument which is based on experiences with visual aids that display only one feature. In working with a device that displays for instance fundamental frequency, one will inevitably be confronted with the child that does produce the required pitch or intonation contour but only at an intensity of 110 dB or with a very bad voice quality. This problem is inherent of mono-feature displays and can only be solved by displaying all relevant information simultaneously.

Assumption 3

The visual aid should display the information in an integrated fashion rather than in a parallel one. The main argument for this assumption is that if the information is displayed in parallel, for instance in different windows on the screen, somewhat like the dials on the dashboard of a car, the subject will probably have great difficulties to monitor all the information simultaneously. Therefore we believe that we should try to develop a system that displays all relevant information in one multi-dimensional form, making use of different independent visual dimensions like form, colour, texture, size etc. A considerable body of research has shown that information presented in independent dimensions (such as those just mentioned) are processed almost in parallel. That is, with two dimensions one can convey almost twice as much information as with one [1].

THE DESIGN OF AN AID

In this way we approach the ideal formulated before, in the sense that all relevant information is displayed in a form that is easily accessible to the eye. But, as mentioned above, this is still only part of the answer, because displaying the information in itself does not say anything about the meaning of the information; it does not show how the information is related to speech. It should be noted though that a straightforward display of speech related information, without any normative function, can be a most useful aid since it can help the deaf child to learn the correspondence between visual and articulatory dimensions.

When those relations are acquired the pupil has learned to control specific aspects of the visual display and thus may be said to have developed some aptitude in controlling articulatory structures relevant for speech production. But the child still does not know the relation between the visual image and speech. So now the pupil must be shown how the different visual dimensions are used in the formation of specific speech acts. Here we can think of giving information about the range of values that are used in speech production, like the range of acceptable intensity, fundamental frequency, F1, F2 etc.

The introduction of normative information

One way to introduce a norm is by using a split screen and showing an example of a model speech act on the upper half, which can be imitated by the pupil on the lower half. The models can either be produced on the spot by the teacher or can be build into the apparatus. Although the teacher-produced model is attractive in the sense that it fits within the usual teacher-pupil interaction, in the case of speech training it has several limitations. For instance in displaying intensity-related aspects of a model the distance between speaker and microphone is crucial, but difficult to control especially if the child uses a headphone-microphone combination. Other aspects like fundamental frequency and timbre are even more problematic in this respect because the ranges of values on these dimensions differ greatly between adults and children. To normalize these differences does not seem easy to accomplish. Therefore at present we concentrate at displaying internally stored criteria of acceptability.

Suppose we display intensity of speech as the brightness of the display on the screen. Then, speaking too softly or too loudly would be indicated by the display becoming almost invisible, respectively unpleasantly bright (with well chosen relation between intensity and brightness). This example shows that with a well chosen visual dimension, the normative information can be presented in a most natural way.

For other aspects of the speech signal this may not be readily feasible. Consider for instance the way normative information is build into the Vowel Corrector developed by Povel [4] and Povel & Wansink [5], an aid for teaching vowels to the deaf. This device displays vowels, either spoken in isolation or in mono-syllabic words, as light spots on a screen such that different vowels project at different areas of the screen. When vowels are entered into this apparatus the spot moves over the screen roughly in accordance with the momentary value of F1 and F2 that respectively determine the X and Y coordinate of the spot. In this mode the device only displays some speech-related aspects of the spectrum, but does not indicate the relation between location of the spot on the screen and speech characteristics. This information is presented by indicating on sheets fixed to the screen the outlines of the areas that correspond to different vowels.

It should be noted that this way of displaying spectral information seems most useful because it combines the two functions mentioned above: it shows relevant parameters of speech in a way that is easily interpretable by the eye, and at the same time it shows the relation between the displayed information and certain speech acts, thus fulfilling its normative function.

MOTIVATION

Besides the two functions just discussed, there is yet another aspect that is probably very important in displaying visual information for speech training purposes. This concerns the desirability that the information be presented in a for the child attractive way, for instance in the form of interesting games, thus maintaining motivation during training. Although we believe that this aspect needs attention, we feel that it is even more important to construct a curriculum incorporating the aid, in which tasks are defined that the child can perform successfully, thus maintaining inherent motivation to learn to speak.

CONCLUSION

To summarize we believe that in displaying visual information for speech training purposes, one should aim at displaying as much relevant information as possible in an integrated visual display using independent visual dimensions that are uniquely related to speech parameters. Further, the device should incorporate norms as to how the different dimensions are used in forming specific speech acts. All these aspects should be part of a training curriculum in which attention is given to factors stimulating motivation. Apart from the specific problems to display the separate parameters, we think that the main challenge will be to combine the different requirements in a workable visual aid.

Currently we are working on two parallel lines. In the first one we develop aids for separate aspects of speech. Here we concentrate on displaying segmental, rather than on supra-segmental information on the basis of the results of the work of Maassen & Povel [3] which has shown that an improvement of intelligibility is mainly found after correcting segmental aspects of speech. In the second line we are building aids that combine different aspects in one complex multidimensional display. Examples of displays will be shown during the presentation.

REFERENCES

[1] Attneave, F. (1959). Applications of information theory to psychology. New York, Holt, Rinehart and Winston.

[2] Lippmann, R.P. (1982). A Review of Research on Speech Training Aids for the Deaf. In: N.J. Lass (Ed.). Speech and Language Advances in Basic Research and Practice. Vol. 7. New York, Academic Press.

[3] Maassen, B., & Povel, D.J. (1985). The Effect of Segmental and Suprasegmental Corrections on the Intelligibility of Deaf Speech. Journal of the Acoustical Society of America, 78, 877-886.

[4] Povel, D.J. (1974). Development of a Vowel Corrector for the Deaf. Psychological Research, 37, 51-70.

[5] Povel, D.J., & Wansink, M. (1986). A Computer Controlled Vowel Corrector for the Hearing Impaired. Journal of Speech and Hearing Research, 29, 99-105.