2442

AUDITORY TRAINING OF THE VOICING DISTINCTION

A STUDY OF CATEGORICAL SPEECH PERCEPTION IN HEARING IMPAIRED CHILDREN

Th.A.M. CRUL

AUDITORY TRAINING OF THE VOICING DISTINCTION

A STUDY OF CATEGORICAL SPEECH PERCEPTION
IN HEARING IMPAIRED CHILDREN

PROMOTORES: PROF.DR. W.J.M. LEVELT

PROF.DR. W.F.B. BRINKMAN

COREFERENT: DR. J.C. MARSHALL

AUDITORY TRAINING OF THE VOICING DISTINCTION

A STUDY OF CATEGORICAL SPEECH PERCEPTION
IN HEARING IMPAIRED CHILDREN

PROEFSCHRIFT
TER VERKRIJGING VAN DE GRAAD VAN DOCTOR
IN DE GENEESKUNDE
AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN
OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. DR. P.G.A.B. WIJDEVELD
VOLGENS BESLUIT VAN HET COLLEGE VAN DECANEN
IN HET OPENBAAR TE VERDEDIGEN
OP VRIJDAG 7 MAART 1980
DES NAMIDDAGS OM 4 UUR

DOOR

THOMAS ANTONIUS MARIE CRUL

GEBOREN TE SAWAH LOENTO (SUMATRA)

Nijmegen 1980
druk: H. Peeters en J. van Kaauwen

aan: Anneke,

Maayke, Michiel en Joris

Een aantal personen hebben een aandeel gehad bij de totstandkoming van dit proefschrift. Van hen wil ik vooral dank betuigen aan:

Ria te Winkel-Slotboom voor haar medewerking aan het gehele onderzoek, de verzorging van het uiterlijk van het proefschrift en het typen van de tekst.

Drs. Ben Maassen voor zijn hulp bij de konstruktie van de stimuli.

Dr. Patrick Huygen voor de statistische hulp bij hoofdstuk 3 en voor het beschikbaar stellen van de omslagtekening.

Ir. Godfried Hombergen en Ir. Theo Zelders voor hun bewerking van het programma van de Pollack-Norman Index.

Ing. Martin Nicolasen voor zijn bijdrage aan de konstruktie van de leerapparatuur.

Leerlingen van de Mgr. Martinus van Beekschool te Nijmegen voor hun aktuele medewerking aan het onderzoek.

Jean Saris, Dr. David Holzman, Drs. Eric Kellerman, Dr. Jan Blom en Dr. Joop Bekkers voor de vertaling van het proefschrift in het Engels.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1. THE AUDITORY APPROACH IN A DIFFERENT PERSPECTIVE.

Speech reaches us for the greater part in encoded form and a specific perceptual mechanism is demanded to convert the complicated acoustic signal into comprehensible form (Liberman et al., 1967a,b; Liberman, 1970, 1973a,b). The greatest complication in the speech code is that the same acoustic parameter in the speech signal provides simultaneous information about consecutive elements (Liberman et al., 1952; Liberman, 1957). Thus it appears practically impossible to divide up a spoken utterance into separate segments such as the phoneme, solely on the basis of acoustic constituents. This means that the listener perceives co-articulated integral units - roughly one syllable in length - in which the acoustic cues are highly related to one another. Research into speech perception is therefore less involved with the abstract phoneme than with the problem as to which acoustic parameters or cues in the complicated speech signal are responsible for its perception. A number of these parameters have been established in fairly recent experiments and can be reproduced in phonetic research. One of them, primarily responsible for the voiced-voiceless contrast, has been used in this research. The fact that speech is in coded form makes its perception a far more efficient matter for the listener, who, as has been shown (Liberman et al., 1967), has too limited a temporal-perceptual solving capacity for speech sounds at his disposal to be able to comprehend them without the code. For this reason Liberman, when discussing the speech decoding processes with the assistance of specific perceptual mechanisms, talks of 'high speed performance with low speed machinery' (1967b). The capacity of always being able to perceive acoustic speech signals as speech is a wonderful fact, despite the very swift spectral changes as important cues, the variation of the signal in the utterance of the same speech sound in different contexts, and the noticeable acoustic differences between the voice characteristics of different speakers (Studdert-Kennedy, 1975). But the very benefits which the speech code grants the normal listener can often be considered as a problem for the auditorily handicapped. How would this complicated process work in the case of someone who, for one reason or another, does not obtain enough information from the spectrum, or who is not in the position to perceive information in encoded form? Does he hear speech in the specific form as the normal listener would, that is to say after transformation of acoustic information into linguistic information, or some other minus variant of speech, possibly even without linguistic-phonetic relevance and only in acoustic fashion?

Frey (1975) assumes that the relative seriousness of the handicap and the distortion which the acoustic signal can undergo in the case of auditory handicap can be graded in terms of the capacity for speech perception ranging from total non-perception of the speech code to something approaching normal speech perception. This gradation is for example reflected in the differences in speech production between on the one hand the speaker who is born deaf and who has learned to speak more or less completely through the proprioceptive and visual modalities and, apart from some suprasegmental cues, without support of auditory feedback, and on the other hand the speech production of someone who is hard of hearing who still perceives sufficient parameters in the acoustic signal for his speech to approach normality. It is difficult to determine exactly how someone with severe hearing problems actually perceives speech sounds. House et al. (1962) have already shown that there is no 'speechlike continuum', that is to say a gradual transition from sound patterns which have nothing to do with speech to an accurate approximation of speech sounds. Liberman et al. (1961) found that a given acoustic variable in a non-speech context can bear some resemblance to the same variable in the speech signal from a physical point of view, but that responses to stimuli containing these variables demonstrate a sharp dichotomy, i.e. either they are perceived as speech or they are not. Where is the boundary between speech and non-speech for an auditorily handicapped person? The all-or-nothing principle in relation to the phonetic perception of the auditorily handicapped person can in theory be seen as less absolute if one works from the assumption that speech perception in the partially deaf or in people with some other form of auditory handicap often takes place in fragmentary form, in less sharply compartmentalized and more fluidly organized categories, than the ones in which speech sounds are normally perceived.

Some evidence that speech is perceived as non-linguistic or as not belonging to the speech mode among people with very severe hearing problems can be deduced from experiments with critical speech perception cues among deaf children and adults (Pickett & Martony, 1970; Martony & Agelfors, 1975; Stark, 1974). Many severely handicapped children do not perceive all the relevant cues but sometimes can rely on the transmission of one or more speech characteristics.

Erber (1972) presented spondae lists for identification under two conditions to children suffering from a loss of hearing that was greater than 90 dB in the frequencies below 2 kHz and without any hearing in the frequencies above this level. In the first condition these spondaes were offered in amplified normal linear form. In the second condition, the words were filtered first through a low-pass-filter, so that the frequencies that were lacking in the hearing of these children were removed from the acoustic signal. Subsequently, keeping the waveform envelope which remained after filtering, he replaced spectral cues with low-frequency noise, so that only a time-intensity pattern for each word remained. The deaf children performed identically in both conditions and the evaluation of speech in the second condition did not even differ qualitatively from that in the first. Thus it appeared that they were unconscious of the fact that they were not perceiving speech in the form of phonemes. Eber's conclusion that these children perceive in a non-phonetic modality and probably only receive crude suprasegmental information, seems to be justified. Schultz and

Kraat (1971) presented a series of consonants for recognition in two different vowel contexts to children who were hard of hearing. After a phoneme- and a distinctive feature-analysis via a signal detection method it appeared that the extent to which the same phoneme or feature was detected was influenced by the two different vowel contexts. They established that the perception of these children is context-dependent, due to too great a dependence on the acoustic signal and an inability to extract identical linguistic information in the form of phonemes from differing acoustic patterns.

Crul (1977a) came to more or less the same conclusion. He taught children who were very hard of hearing to discriminate initial voiceless stop consonants in a fixed vowel context up to a certain criterion and afterwards presented these children with the same consonants in another vowel context in order to teach them to discriminate. Comparison of the results of both successive learning experiments showed that no transfer of information had taken place from the stimuli in the first to the second context, and that response relationships in the first experiment were quantitatively different from those in the second. Here too the question was raised as to whether extraction of phonemes from the acoustic signal was a reality for these children or whether they had not learned to distinguish one acoustic pattern from another as a holistic Gestalt. This suspicion was strengthened when it appeared from the results of a long term retention study, using identical stimuli to those in the second learning study, that the answer patterns in this case did indeed agree with those of the second learning study. That is to say that the children responded to the consonants in two different contexts in systematically different ways.

The question whether disturbed auditory perception of speech can be improved is fundamental to every form of auditory training. Ever since this has been applied reasonably systematically (see Wedenberg, 1951; Watts, 1969, for a historical review) teachers propagating the speech method in the teaching of the deaf and the hard of hearing, and subsequently many concerned with speech and learning disabilities, have considered this training as a primary goal (see Winitz, 1969 for the former and Oakland and Williams, 1971 for the latter). Many of these methods combine analytical ear training (e.g. by practising with non-speech sounds or 'isolated' phonemes) with a more holistic approach with the help of words, and they often claim that, through the training of all kinds of auditory subfunctions, they can bring about a gradual improvement in perception, e.g. by means of crude non-speech sound training up to and including the training in fine speech discrimination (cf. Zigmund & Cicci, 1968; Pollack, 1970; Jensen & Jussen, 1970; Heasly, 1974). The evaluation of these programs is often merely quantative, without their providing any insight into the fundamental learning process.

As for the question of what effect auditory training has it seems of primary importance to find out at what level the disability lies, i.e. in the transmission system for acoustic information or in the specific mechanism that uses this information and converts it into a linguistic message. Conventional training methods seem useful in the first case, though this is questionable in the second. It is for instance very uncertain whether and in what way positive transfer of non-linguistic information, obtained through more or less arbitrary analytical non-speech programs, to specific speech perception actually takes place (Ling, 1978). Recent research shows that an acoustic speech cue carries relevant information only as an integral constituent of the speech signal and has no independent linguistic significance; it may well even lead to a fundamentally different process of perception (non-categorical). Such a cue appears then to be phonetically relevant within the speech signal but only of acoustic significance outside the signal (Liberman et al., 1961, 1967a,b; Lisker et al., 1962; Eimas, 1963; Mattingly et al., 1971). The acoustic and phonetic stimuli would appear to be processed in a likewise fashion through various perceptual mechanisms. There is however a plausible alternative theory that bases the difference between the acoustic and linguistic mode of auditory processing on a psychophysical model instead of a neurolinguistic explanation of the perception of non-speech versus speech stimuli. This theory provides an explanation for a variety of characteristic aspects of perception that are considered as unique to speech perception, on the basis of various perceptual strategies that appear to depend on the series of stimuli and the nature of the task and that make use of various memory modes (Ades, 1977).

The holistic side of auditory training also has its critics. Without the possibility of converting into phonemes the complex acoustic speech signal on the basis of the underlying encoded cues and through a restructuring process - a process that every normal child learns naturally - we should be obliged, according to Liberman et al. (1967), to present all our different ideas by means of sounds that also differ uniquely from one another; one sound pattern for each idea. This implies as many thousands of sound patterns as there are ideas, which would demand from the speaker-listener an impossible capacity. On the contrary, the efficiency of the speech code consists theoretically in the fact that roughly 40 phonemes can be used to represent an infinite number of ideas. The objection to a quantitive evaluation of a method is that an improvement or a falling-off in the desired result can only be expressed quantitively while there is no possibility of analysing the underlying processes. This seems to be the case for a holistic approach.

Rees (1973) warns against persistent claims that a long series of disabilities in auditory functions that have not been subjected to valid research would be the basis of language, speech and learning disabilities, as is stated in survey articles by, inter alia, Weiner (1967) and Flower (1965). She also takes up a critical attitude towards the number of test procedures based on these claims, designed to measure auditory capacities for clinical or educational diagnoses. She is also critical of auditory training programs deriving from these claims which maintain that they can correct speech behaviour and learning disabilities. She posits that recent data on speech perception in many cases tend rather to argue for the existence of a defect in the linguistic capacity and not in the purely auditory one. This direct attack on all kinds of auditory function tests and function training methods is fully understandable in the light of the above.

In the research reported on here the effect of auditory training on the voiced-voiceless phoneme contrast via a controlled speech cue which is part of a normally spoken word is examined in several auditorily handicapped children. What we mean by auditorily handicapped children is the following: children who either cannot perceive this cue and its linguistic effect, i.e. the contrast between the feature 'voiceless' and the feature 'voiced' in initial consonants or only do so in abnormal fashion. Malfunctioning perception may be due to faulty transmission of the acoustic speech signal, for example in the case of a hearing disability, or due to a defect in the speech decoding mechanism, such as we find in dysphasia. We chose the auditory approach method in view of the fact that hearing is the most natural way of decoding the complex speech signal, and because in the past it has appeared that efficient decoding is hardly or not at all possible via other sense modalities (cf. Potter, Kopp & Green, 1947).

Liberman (in Stark, 1974) suggests the use of controlled cues in the speech signal in the training of auditorily handicapped persons and recommends that one should make use of the knowledge of the relative importance of the differing constituents in speech. He supposes that once a person with this kind of handicap gets used to a number of synthetically processed relevant cues on the basis of a reinforced learning process, it will be easier for him later on to abstract these cues from a speech signal which has not been clearly received.

Giving training in one contrast represented here by the features [+voice] versus [-voice] in prevocalic stop consonants, has the advantage that the listener concentrates on a limited set of cues. Consequently it may be expected that the quantitative learning effects and the measure of generalization to contexts which have not been of issue during learning will be greater than if the set to be learned had been more varied (Bennet & Ling, 1977).

We chose Voice-Onset-Time (VOT) as the speech cue responsible for the perception of the voicing distinction. As an integral constituent of a word this cue is technically manipulable and thus offers one the opportunity to successively serve as critical stimulus in diagnostic, training and effect of training research.

Given the fact that the linguistic effect of VOT in a speech signal, i.e. the recognition that a consonant is voiced or voiceless, is normally perceived in categorical fashion (Liberman et al., 1958, 1961a,b; Lisker, 1970; Lisker et al., 1962), the paradigm of categorical speech perception can be followed in the pre- and posttraining phases of the learning study. This provides the opportunity for accurate measurements of essential changes in the perception of the voicing contrast in relation to the normal values which can be obtained.

In a wider context the effect of an improvement in perception can be measured by means of lists of normally spoken words, which contain the distinctive features being examined, again in prevocalic stop consonants but in a different vowel context, and which can be offered for comparison prior to and after training.

In short the aim of this research is to see whether the abnormal perception of a distinctive time cue in the speech signal can be improved by a learning experiment and whether this improvement applies equally to situations where this cue causes the distinctive effect in other contexts in natural speech; in other words, whether a small but fundamental improvement will create a large effect.

1.2. SURVEY OF EXPERIMENTAL PROCEDURE.

1.2.1. Construction of stimuli.

In the first phase of the research a stimulus continuum was constructed under computer control, with a VOT variable as critical acoustic parameter, consisting of two related bilabials /b-p/ in the context [bɑk-pɑk] and the related alveolar consonants /d-t/ in the context [dɑk-tɑk], both pairs being spoken naturally. Randomized identification and discrimination lists were produced of both sets of stimuli under computer control and recorded on tape. These were offered to a group of male and female subjects aged between 4 and 40, the hypothesis being that the stimuli would be perceived according to the assumptions of categorical speech perception. The boundary values could be determined for each individual identification curve obtained. Data reduction was applied by eliminating the results of subjects who fell outside a given criterion with regard to these values and by testing the remaining data for differences in sex and age.

These procedures oughted to show whether individual data could be pooled. From these pooled data the normal identification and discrimination functions of the /b-p/ and /d-t/ continuum could be subsequently determined. The pooled boundary values of the individual identification functions yielded four reference parameters per stimulus series with were used in later learning study as norms.

1.2.2. Auditory training.

In the second phase a group of auditorily handicapped children, part of a population of a school for children with speech and hearing disabilities, was selected on the basis of inadequate perception of one or both perceptual distinctions. The auditory handicap could consist of a hearing disability or a disability in a higher level auditory perception process.

Two children with relatively slight hearing deficiencies, two with very serious hearing deficiencies and three children with speech disabilities, all with various aberrations in their perception functions of one or both VOT series, were included in the experimental learning group, while four children with various forms of hearing deficiency functioned as the control group. The experimental group was examined in order to determine whether the aberrant perception of voicing features in initial consonants could be fundamentally changed for the better and what effect such a fundamental change would occasion in the perception of voiceless and voiced initial consonants in contexts not included in the learning. The control group served to show whether the same learning and generalization effects could be established without the specific learning procedure.

1.2.3. Transfer of learning.

The successive stages of the phase 3 procedure, the learning experiment, are shown in the flow-diagram in figure 1.

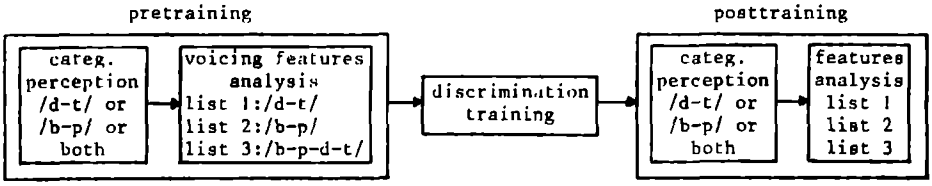

Figure 1: Flow-diagram of successive steps in the learning program.

Apart from the collecting of data about the way in which the two VOT continua were perceived in the pre- and posttraining phases, the experimental subjects were analyzed both before and after training as to their perception of the features of voicing in initial consonants in various vowel contexts. This was done with three different word lists with multiple choice items whose stimulus word was spoken normally and all of which began with a prevocalic voiced or voiceless stop consonant. The first list contained alveolar voiced contrasts /d-t/, the second list bilabial voiced contrasts /b-p/ and the third list consisted of not only voicing contrasts, but also contrasts in the place of articulation, that is to say bilabial versus alveolar /d-t-b-p/.
Only the perception of voicing features was included in the analyses. After obtaining the data from the pretraining research there followed operant discrimination training in which stimuli from the VOT continuum were used which were selected on the basis of the perceptual aberrations of the individual subject. After reaching a certain criterion the identification lists were again presented with the VOT dimension perceived earlier as aberrant, in addition to the three word lists. The changes in the perceptual functions of the VOT continuum and in the capacity to perceive the voicing features in the three lists could then be interpreted in the light of the questions posited and the hypotheses.

In the pretraining phase the control group was presented with the categorical perception tests and the /d-t/ and the /b-p/ lists. They did not receive any training but were again presented with these word lists after roughly the same interval as the experimental subjects. In their case too the differences in feature extraction between the first and second presentation were interpreted and compared to the results of the experimental subjects.

1.3. PROBLEMS.

Before going on to discuss a number of questions relating to the possibility of bringing about a fundamental improvement in the perception of feature contrasts we would like to discuss some problems that are directly connected with this.

Question 1: *Is it possible to improve aberrant or non-categorical perception of the normally categorically perceived feature contrast 'voiced-voiceless' in certain pre-vocalic word-initial stop consonants by means of specific auditory training?*
Certain theoretical implications in connection with the possible existence of congenital and pre-determined categorical speech perception and with learning to perceive

speech sounds through experience are dealt with in Chapter 2.

Question 2: *Should there be improvement in the categorical perception of certain pre-vocalic word-initial stop consonants with a voicing contrast, would there also be transfer of information to stop consonants that share the same position within the syllable and the same place of articulation and the voicing feature with the trained stimuli, but are in a different vowel context?*
This concerns the question whether the change for the better in the perception of the features [+voice] and [-voice] in certain initial consonants also exerts a positive influence on voiced and voiceless initial consonants in a varied vowel context (e.g. /pɑk/→/pe:n/). Closely releated to this question is the problem concerning the presence or absence of invariances in speech perception. The following chapter contains a survey of this question.

Question 3: *Should there be an improvement of the categorical perception of certain pre-vocalic word-initial stop consonants with a voicing contrast, would there also be transfer of information to stop consonants that share the position in the syllable and the voicing feature with the training stimuli, but that are in a different vowel context and also differ as to the feature of place of articulation?*
We are concerned here with the question as to whether a feature (e.g. [-voice]) is perceived independently of the place of articulation, thus resulting in transfer of information to an initial consonant, whether in the same vowel context or not, that differs as to place of articulation (e.g. /pɑk/→/te:k/). Therefore we are concerned here with non-place-bound independently operating features that can be perceived outside the phoneme context. Chapter 2 contains some observations as to whether or not features operate independently.

CHAPTER 2

VOICE-ONSET-TIME, A 'TRAINABLE' CUE?

This chapter deals with a number of theoretical and practical implications in connection with the experimental use of VOT as a fundamental speech cue in a perceptual learning study. Although in recent phonetic linguistic research VOT appears to be one of the most manipulated dimensions in the speech signal, there are essential contradictions which have so far not been resolved.

The question as to whether VOT perception can be trained and whether, in the case of learning, positive transfer to other contexts can take place can only be examined in the light of recent theories and concrete findings.

2.1. VOICE-ONSET-TIME.

Voicing is one of the perceptual dimensions by which consonants can be distinguished. The class of stop consonants can be grouped together in cognate pairs which have the same place and manner of articulation but contrast in terms of voicing.

This research deals with the bilabial pair /b-p/ and the alveolar pair /d-t/ in word-initial, prevocalic positions.

The articulation of an initial stop consonant demands complete closure of the vocal tract, causing a build-up of air pressure, which is then released in a burst, and the articulations then move towards the positions appropriate for the following vowel (Massaro, 1975).

The perception of the voicing contrast in Dutch speech is to an important extent caused by the presence or absence of vocal cord vibration during a period of closure preceding or simultaneous with the release or following the release. Voiceless stop consonants are usually followed after the burst release by a period of aspiration before the vocal cords begin to vibrate in the vowel portion (the voicing lag). In the case of voiced initial stops in Dutch speech the vocal cords often begin to vibrate before the burst release (prevoicing or voicing lead) and remain like this without a period of aspiration in the vowel portion (Lisker & Abramson, 1964; Slis & Cohen, 1969a,b). The acoustic features for the perception of voicing are then the differences in voice onset time (VOT) in relation to the burst release and the presence or absence of aspiration. Further perceptual dimensions relating to voicing contrast have been described by Slis and Cohen (1969a,b), Stevens and Klatt (1974) and Klatt (1975). However, it appears that VOT is one acoustic dimension which can suffice to bring about perception of voicing on its own (Liberman et al., 1958; Lisker & Abramson, 1964; Lisker, 1970, 1978; Lisker et al., 1977; Slis & Cohen, 1969a,b), so that aspiration can be considered as a redundant cue. Therefore VOT can be defined briefly as the time between the release burst and the onset of laryngeal pulsing and is seen in this investigation as a sufficiently reliable cue for bringing about the perceptual distinction between a voiced and voiceless initial consonant.

In experimental phonetic research the VOT dimension in stops is represented as an acoustic continuum expressed in msec. The beginning of the burst release on this continuum receives a VOT value of 0 msec. A negative value is assigned if vocal cord vibrations occurs while the oral cavity is still closed prior to the burst release (voice lead). The positive value points to the onset of vocal cord vibration after the burst release (voice lag).

Despite the fact that VOT can be represented in the speech signal as an acoustic continuum it appears from identification experiments with separate stimuli taken from the continuum that their perception by children and adults proceeds categorically (Liberman et al., 1957, 1958, 1961a,b; Eimas et al., 1971; Eimas, 1974, 1975; Zlatin & Koenigsknecht, 1976; but see also Ades, 1977). This means that no gradual transition from one phoneme to the other is perceived, corresponding to a change along the acoustic continuum, but rather an abrupt transition to the contrasting category. It would thus appear that, on the basis of VOT variations alone, speech stimuli are grouped perceptually in phoneme categories with clear phoneme boundaries. This compartmentalization into clear categories with a boundary area on the VOT continuum is specific to categorical perception.

A second specific property of VOT stimuli in the speech signal becomes apparent from discrimination experiments with two stimuli from the same VOT continuum. If the acoustic differences in VOT between stimuli are held constant and if these are offered in pairs for discrimination, it appears that differently voiced phonemes are perceived if two stimuli with a given VOT value lie on opposite sides of the phoneme boundary (between-category discrimination). But discrimination is a hardly better than chance occurrence if two stimuli have to be discriminated which vary in VOT within the same category (within-category discrimination). The discrimination curve reveals a trough for discrimination within categories and a peak for discrimination between categories. This is in contradiction to the perception of less discrete sounds such as vowels (Fry, 1962; Stevens et al., 1969) and non-speech sounds (Eimas, 1963), which are monotonely discriminated without peaks or troughs in the perception curve.

As the third specific property it is postulated that the discrimination results can be predicted with a great degree of reliability from the identification performances (Liberman et al., 1957, 1958, 1961a). This is considered essential for categorical perception (cf. Studdert-Kennedy et al., 1970).

Earlier we referred to the alternative theory presented by Ades (1977), on the basis of which the characteristic properties of categorical perception are not explained in speech specific terms, but in psychophysical ones.

To sum up it can be said that the listener perceives categorically various stimuli as voiced or voiceless phonemes on the basis of VOT variations in the speech signal, without his being able to perceive the acoustic differences.

The three fundamental properties concerning the categorical perception of stop consonants will be made further use of in the construction of experimental stimuli in this research in order to obtain valid VOT continua.

Of indirect relevance for this research is the assumption that the distinction between 'voiced' and 'voiceless' is universal. In accordance with the observation by Jacobson (1968) as to the universality of these features, well-known cross-language studies by Lisker and Abramson (1964, 1967) have proved empirically that the distinction between 'voiced' and 'voiceless' on the basis of the VOT speech cue has been found in many languages and that is why these features can be counted as universal speech sounds in the perceptual-productive sense.

Another property, finally, of direct significance for auditory training and speech therapy has recently been actually demonstrated. This is the close relationship which exists between perceptual and productive VOT values of voiced and voiceless stop consonants by means of which the functional link between perception and production of speech sounds again appears to be confirmed (Zlatin, 1974; Zlatin & Koenigsknecht, 1975, 1976; Bennet & Ling, 1977).

2.2. AUDITORY TRAINING AND THE PROBLEM OF INVARIANCE IN THE SPEECH SIGNAL.

Although the relatively recent study of the speech signal and perception of speech has led to a better understanding of critical acoustic cues, it is also quite clear that this signal is a complicated pattern of very variable overlapping acoustic segments without an isomorphous relationship between various segments and the phonemes perceived.

This problem is important in the application of auditory training with the aid of a certain property in the speech signal. Learning, as is the case in this investigation, to perceive stop consonants as the same speech sound despite variation in the acoustic contexts presupposes the existence of recognizable invariant cues in the signal. Such invariances can be defined as 'acoustic cues with accompany a particular phoneme in any vowel environment' or as 'acoustic features which are present whenever a particular consonant phoneme is produced during speech and which provide information about the phonetic identity of the consonant' (Cole & Scott, 1974a,b).

One of the central problems in research into speech perception appears to be, however, that no invariant cues are found as a result of the movement of the articulatory apparatus that produces the sounds (Liberman et al., 1967). Early experiments by the Haskins group have shown that,in the case of consonants, there is no 'simple one to one correspondence' between perceived units and the acoustic structure of the signal and that the stop consonants are perceived as context dependent units because a segment of the acoustic signal contains information about different phonetic segments. It also appeared that acoustically differing signals in varying contexts could lead to the same phonetic perception, and conversely that identical acoustic segments could lead to different phonetic perception (Liberman et al., 1967).

Opinion as to which units should be considered responsible for normal speech perception is somewhat divided and covers the whole range from sub-phonemic acoustic information up to and including whole parts of sentences (Studdert-Kennedy, 1970; Lehiste, 1972; Shankweiler et al., 1975).
Cole and Scott (1974a,b) have tried to demonstrate that the perception of speech is based to an important extent on the discovery of acoustic invariances. However, the invariances in stop consonants which they propose also appear to contain information about neighbouring segments and are thus not accepted as unique speech cues (cf. Darwin, 1976).
The following can be said about VOT as a stimulus in research: it can be considered as a relatively stable bearer of information in initial stop consonants in the speech signal. Nevertheless this speech cue is not entirely invariant either and it appears that there is a certain speaker- and context-dependency. VOT boundaries can, for instance, vary from language to language and from speaker to speaker; they are dependent on the place of articulation of the consonant, the following vowel, the pitch of the speaker, the speed of articulation and stress (cf. Lisker & Abramson, 1967; Lisker, 1970, 1978; Klatt, 1975; Lisker et al., 1977).

The existence of invariant cues in the speech signal would be a plausible explanation for the perception of speech. In terms of a filter theory unusable information would, for example, be rejected and critical invariant cues from the speech would be retained. A passive speech perception model of this kinds would, however, no longer be considered acceptable and in place of it the perception of speech is seen as an active process in which the phonetic analysis of the speech signal, in different successive phases, is realized as an integration of acoustic information with the linguistic backgroundknowledge which the listener has at his disposal.
Proponents of active perception models, such as the Motor Theory of Speech Perception put forward by the Haskins group(Liberman et al., 1967) and the similar Analysis of Synthesis Model of Stevens and House (1972, see also Stevens, 1972), posit a hierarchical sequence of information processing which is realized on different levels and by means of which the acoustic signal is converted into linguistic reality (see also Pisoni & Sawusch, 1975).
However, background information is based on information from the speech producing apparatus or on what are known as inbuilt phonological rules.
In this present research in which only a limited aspect of speech perception is being studied (i.e. an auditory phonetic learning process with abstraction from phonological, syntactic and semantic levels of the perceptual hierarchy) postulation of background knowledge based on the speech of the listener himself is a dubious source of information. Speech of children with speech disabilities and who have auditory perception problems and of children with hearing disabilities (both groups here being considered as auditory handicapped) is by definition aberrant and offers virtually no guarantee that it will function as a reference model in a perceptual process.The Motor Theory of speech perception and related theories do not offer any sufficient basis on which to

justify the auditory approach for auditorily handicapped children. What is more, we may refer to Lenneberg's well-known example (Lenneberg, 1962) of patients with speech disabilities who have never learned to speak, nevertheless appearing to be able to learn to understand speech (see also McNeilage et al., 1967; Fourcin, 1975).

A better starting point on which to base a purely auditory approach would seem to be the recent perception model proposed by Eimas (1974) and Cutting and Eimas (1975a) and the first two highly similar stages of the perception model of Pisoni and Sawusch (Pisoni & Sawusch, 1975).

It is suggested that in the first stage of this model a system of individual auditory property detectors is operant which are sensitive to various types of specific acoustic information in the speech signal and which mediate the perception of various phonetic distinctions in a second stage. These detection mechanisms work at a lower level.

In the second stage a set of decision rules are used to classify the multiple auditory patterns extracted from the first stage into phonetically distinctive characteristics on the basis of feature detectors. These detectors function at a higher level. The decision rules are not necessarily related to the speech production of the listener (see also Fourcin, 1975). The individual features can eventually be recombined in order to form the discrete phonetic segments.

A disadvantage of a property and feature detection model is that it again works on the assumption of the existence of invariant cues or stimulus properties in the speech signal to which the detectors are particularly sensitive (Stevens, 1975), while in fact this ought to be established by experiment.

The hierarchical sequence of an auditory and phonetic processing of the speech signal presupposes that no speech segments will be identified at the auditory level (with the auditory short-term memory operant here) and that a pattern of auditory cues is identified as an abstract phonetic segment on the phonetic level (with the phonetic memory that is by nature categorical) without the cues on which the identification is based being retained (Studdert-Kennedy, 1970; Pisoni & Sawusch, 1975).

In the present research there is no a priori knowledge as to whether auditory training with a temporal stimulus property (VOT) as an integral part of a syllable will operate in the case of a certain auditorily handicapped subject on an auditory, lower level or a phonetic, higher level. Learning on a purely auditory, lower level will result in an improvement of the perception of auditory differences without a resultant linguistically relevant effect.

We assume that one can speak of linguistically relevant learning when the improved perception of the feature contrast 'voiced-voiceless' in initial stop consonants will also imply an improved perception of the same consonants in different vowel contexts or of initial consonants that occur in a different vowel context and share the same voicing feature, but are produced in a different place of articulation. In this case one could accept that the perceived features of 'voiced' and 'voiceless' operate as context-independent units.

The following findings and arguments can be used to support the existence of such units.

The speech perception experiment, in which the phoneme boundary is determined with the aid of a variable acoustic cue along a speech continuum, can be considered as a search for the boundary between two linguistic features that are situated on either side of that boundary and differ as to one single articulatory feature (Studdert-Kennedy, 1970).

Miller's and Nicely's analysis is well-known (1955). It was carried out on the confusions that occur in the perception of English consonants in various conditions and from this it appeared that the perception of a phoneme does not take place according to an all-or-nothing principle, but that mistakes can be made for each individual feature over a number of phonemes with the same feature. More recent analyses applying other methods show that in the case of people with hearing disabilities the confusion patterns vary according to the degree and the form of the loss of hearing. Success in the perception of certain distinctive features is dependent on this hearing loss. It appears that the distinction 'voiced-voiceless' is extremely resistant to noise or transmission problems resulting from a loss of hearing. These studies show that the distinctive feature is also a reality for the person with hearing disabilities (Owens et al., 1972; Danhauer & Singh, 1975; Bilger & Wang, 1976; Wang et al., 1978).

From a feature analysis concerning the memorising of phonemes, Wickelgrenn (1966,1969) concluded that vowels and consonants are not represented in the short-term memory as units but as a set of distinctive features, each of which can be forgotten relatively independently.

Further evidence for the existence of individual features can be shown by means of the analysis of systematic articulatory mistakes in speech production (Fromkin, 1973) or on the basis of the research into cerebral lateralisation and dominance in the perception of specific speech sounds or their underlying features (Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970).

The most recent indications for the existence of independent phonetic units in the speech signal to which specific detectors of the perception mechanism are tuned derive from the results of perceptual adaption experiments (Cooper, 1975). The adaptation procedure is based on the assumption that a frequently repeated exposure to a certain stimulus weakens the subsequent response of the detectors that are especially sensitive to this. This weakening occasions a disturbance in the subsequent perception of each stimulus that normally would have been able to activate these detectors.

In the first experiments, in which the adaptation procedure with speech stimuli was applied, a crossed adaptation effect was observed (Eimas & Corbit, 1973; Eimas et al., 1973). The frequent repetition of a stimulus selected from the extreme end of a /p-b/ VOT series caused a systematic shift of the phoneme boundary (whose position had been determined earlier) at a subsequent identification of the whole series. This shift worked to the disadvantage of the category from which the adapting stimulus was derived. But frequently repeated presentation of an extreme stimulus from a /p-b/ VOT series also brought with it a corresponding shift in the phoneme boundary of a /d-t/ VOT continuum, the stimuli from which differed from the first only in the place of

articulation. The effect occurred whether the adapting stimulus derived from the voiceless or from the voiced category. An identical shift in phoneme boundary was established if the discrimination tests of the relevant VOT series were offered.
The appeareance of a crossed phoneme boundary shift in the two VOT series suggests most strongly that the adaptation exerts a selective influence on a mechanism that extracts information concerning the voicing property of consonants and not on a mechanism that processes the consonants as a total unit. What is more it suggests the presence of a common property between the adapting and the adapted stimuli from both VOT series which differ only in the feature 'place of articulation'. This common property is the feature of voicing which was realised acoustically by variations in VOT. The same results were also found in adaptation experiments in which the adapting and adapted stimuli shared the feature of voicing and differed in the feature of place of articulation.

The most important conclusion to be drawn from this results for the present learning research is that the feature of voicing can be influenced independently of the place of articulation and vice versa, which argues for the existence of a mechanism which extracts phonetic feature information; this cannot easily be explained by means of any invariant acoustic property.

The results of the selective adaptation with stimuli based on VOT argue for the existence of a twofold detector mechanism: one that is specifically tuned to the stimulus properties of the voiced stimuli and one that is tuned to the properties of the voiceless stimuli (Eimas et al., 1973; Eimas, 1974; Cutting & Eimas, 1975).

Theories about the independent existence of phonetic features, which originate from facts from adaptation experiments, were still further strengthened when it also appeared that the adaptation effect did not limit itself to any single sense modality.

In the meantime further indications have been found for the existence of more cross-modality adaptation effects. A feature specific effect on speech production can be brought about with an adapting auditory stimulus (Cooper & Lauritsen, 1974; Cooper & Nager, 1975) and vice versa (Cooper et al., 1975, 1976).

This perceptuo-motor adaptation again poses the problem of the common processor which is said to be the basis of speech perception and speech production.

The discovery of adaptation effects in an audio-visual procedure (Cooper, 1975) raises suggestions for the part played by the independent phonetic features in the complex linguistic process of reading.

Summary: In the speech signal there are no invariant cues to be found which are unique for the effectuation of constancy in speech perception.

This creates a problem as far as explaining the transferable perception learning of speech sounds (i.e. learning to perceive the same phoneme in every possible context).

The link between speech perception and speech production which is postulated in an active speech perception model in order to explain this problem does not seem to be very suitable as an explanation of a perceptual learning process in the case of children with auditory handicaps and speech aberrations. This is why this research bases itself on an auditory, but not necessarily speech bound, learning process on two levels. A lower auditory process in which certain stimulus properties are abstracted from the acoustic signal by property detectors and a higher auditory process in which the information from the lower process is converted into phonetically and linguistically relevant information (i.e. distinctive features) by means of special feature detectors.

Enough evidence is available in the literature regarding the existence of distinctive features and the feature detectors which are actively in operation.

A feature analysis will be used in this research in order to examine a generalizing effect in speech perception as the result of the improvement in the perception of a fundamental speech cue.

2.3. CAN SPEECH PERCEPTION BE CHANGED FUNDAMENTALLY?

There is a degree of uncertainty as to whether the perceptual parameters of the speech cues are subject to selective modifications under the influence of certain linguistic environmental experiences during the early development of the child. This uncertainty is caused by relatively recent attempts to elicit linguistically relevant responses to speech cues from infants. The results have led to the postulation of a nativistic theory which assumes that human speech perception is genetically determined and is never or hardly ever influenced by experience. Readers will be familiar with the stir created by the article by Eimas et al. (1971, see also Cutting & Eimas, 1975).

By manipulating relevant cues in synthetic speech stimuli and by using a conditioned sucking reflex they showed that children aged between 1 and 4 months can discriminate changes in the relevant cue for voicing and what is more that they can do this in a linguistically relevant way. In view of the fact that at this age the child would not have gathered enough linguistic experience to explain these adequate responses, it was assumed that human beings are born with the capacity to perceive speech which belongs to the 'biological make-up',and that the child is born 'pre-wired to perceive speech'. These relatively provocative statements elicited many reactions (cf. Butterfield & Cairns, 1974; Morse, 1974). What was particularly rejected was the assumption that the perception of speech cues in the case of infants proceeds in a linguistically relevant way and that the speech perception mechanism is complete at birth. Sufficient arguments could be adduced to show that the discrimination of the relevant speech cues does not concern a phonetic categorisation of the stimuli, that is to say a linguistic mode of auditory perception, but is rather a question of the detection of relative differences between two acoustic patterns, that is to say a lower level of auditory perception. The discrimination and identification experiments with animals using natural and synthetic speech stimuli argue against a theory that speech perception is based on an inborn species-specific mechanism.

In this context it has been shown that a non-human primate like the rhesus-monkey can perceive a bilabial VOT contrast while its performance in that kind of experiment strongly resembles that of an adult human being (Waters & Wilson, 1976). Even an animal phylogenetically so far removed from the primates as the chinchilla appears to be able to discriminate and identify the voiced-voiceless distinction in initial stop consonants, as has been proved by means of learning experiments with synthetic speech stimuli (Kuhl & Miller, 1975, 1978).

In the meantime the findings of a number of researchers with similar results to Eimas et al. (1971), have provisionally strengthened the nativistic theory of speech perception or in any case have suggested something of that order.

During her research into sound reactions in the case of newly-born children Eisenberg (1976) found that newly-born children can discriminate sounds on the basis of diverse acoustic variables, a number of which are fundamental for the decoding of speech signals. It was also found that infants responded differently to frequencies which either are or are not important for speech. Speech frequencies elicit a quiet response, other frequencies often result in startle or defense responses. These differences in response are present before birth and suggest a specific aptitude for learning to speak from the moment of birth.

In an early article in which they plead for the existence of feature detectors for speech perception, Abs and Sussman (1971) posit that the underlying neural system which is sensitive to certain acoustic features from the speech signal is innate and can be seen as analogous to the species-specific detector systems which have been found in animals for the purposes of communication (cf. Capranica et al., 1965).

Moffit (1971) found that children aged between 20 to 24 weeks were able to discriminate between synthetic stimuli which varied along an acoustic dimension which signals the feature 'place of articulation'. For this purpose he made use of a cardiac response paradigm. He also maintains that linguistic perceptual capacities are present in the very early development of the child and that infants come into the world with certain knowledge of the phonetic structure of language already available.

Using the sucking reflex mentioned previously, Morse (1972) found that children between 40 and 54 days old could discriminate the place of articulation feature. He claims that production is not a necessary condition for the perception of speech. The perception of the speech code does not have to be learned but is physiologically endowed and not ontogenetically acquired.

Trehub and Rabinowitz (1972) also made use of the sucking reflex in order to investigate the discrimination responses of children aged between 4 and 17 weeks using synthetic or natural voicing contrasts. These responses were indeed found, but in this case uncertainty as to their linguistic relevance was expressed by the researchers.

Fodor et al. (1975) record positive results in an investigation in which the stimuli used can be said to belong to the most complex encoded speech sounds. They conditioned children of 14 to 18 weeks to react to consonants differing in place of articulation and in vowel context by movement of the head. Their results lead them to suppose that these young children have access to the same decoding mechanisms which bring about perceptual constancy in adult speech perception and that these mechanisms are available to them without experience in the production of articulated speech or of the features of the language.

Miller and Morse (1976), using synthetic stimuli, found from the cardiac response that children aged between 3 and 4 months could discriminate the place of articulation feature categorically. They speculate that auditory feature detectors mediate the perception of speech-relevant cues and that these may well be characteristic of the perceptual system of primates.

In contradistinction to nativistic theories of speech perception there are theories in which innate mechanisms for the decoding of speech signals are assumed to operate in conjunction with the learning factor and linguistic experience. Linguistic experience includes hearing one's own speech sounds (as in a closed loop feedback mechanism, cf. Fairbanks, 1954) as well as listening to the speech of another speaker.

Stevens and House (1972; see also Stevens, 1975) maintain that the sensitivity of the auditory system to features and segments used in speech must be partly innate. However, categories appearing in speech are learned because the human being is both speaker and listener at the same time. It is through the link that exists between listening and speaking that he becomes aware of the transformations necessary for the conversion of auditory patterns into productive articulation and vice versa. Thus the child learns the underlying auditory patterns through the use of his articulatory mechanism.

There is considerable similarity between this theory and the earlier Motor Theory of speech perception proposed by the Haskins group. In their attempt to explain the problems of coarticulation effects in the perception of speech supporters of the Motor Theory assume that the perception of speech is more closely related to articulatory patterns than to acoustic ones: in other words speech is perceived by reference to articulation (Liberman, 1957; Liberman et al., 1961b, 1967; Studdert-Kennedy et al., 1970).

How the categories in which we speak or perceive are learned is suggested by Liberman (1957) in an alternative learning theory, the theory of acquired similarity and acquired distinctiveness. It had been established that in discriminating stimuli from a speech continuum with equal acoustic differences, the listener could distinguish more easily between sounds to which he would normally attach different phoneme labels than between sounds that he would normally count as belonging to the same phoneme category. The theory of acquired similarity maintains that for a naive listener all stimuli in such a continuum will be discriminable to the same extent initially, but as a result of linguistic experience in the form of discrete articulatory categories, the sensitivity to differences within the phoneme category declines while sensitivity to differences between the categories remains. This theory does not find much support (Mattingly et al., 1971) contrary to the theory of acquired distinctiveness which has been put forward particularly as an explanation for the origin of a categorical perception of phonemes. The latter assumes that for a naive listener all stimuli are initially discriminable to the same small extent, but that in the course of time discrimination is selectively sharpened as a result of greater experience of speech.

In the course of this experience with speech, a speaker who is at the same a listener, learns to connect speech sounds with their corresponding articulations. After a while these articulatory movements and their sensory feedback become part of a perceptual process and they function as a mediator between the acoustic stimulus and the ultimate perception of the phonemes. If certain acoustic cues, which lie in different positions on a single continuum, are part of the speech signal of essentially discontinuous articulations, their perception also becomes discontinuous or categorical (Liberman et al., 1961b).

Examples of the influence of an external linguistic environment on the development of speech perception and production are to be found in cross-language studies of Lisker and Abramson (1964). They established that the position of phoneme boundaries on a VOT continuum like the number of phoneme categories perceived and produced on the basis of the VOT dimension, is language dependent. Despite their postulation of the universality of certain phonemes, it appears that the acoustical variability and the number of related phonemes is evidently influenced by the linguistic environment.

Another important argument for the influence of environments on speech perception and production comes from studies about phoneme distinctions which are perceived by listeners in a given speech environment where these distinctions belong to the phonological repertoire and are not perceived by listeners from other linguistic environments where these are lacking. Miyawaki et al. (1971a,b) describe the phenomenon of the English phonological contrast /r-l/ not being perceived by Japanese adults, who do not produce this contrast in speech. This contrast is not perceived by them as categorical, but as a non-speech signal. But Japanese who spent part of their childhood in an English environment do perceive the /r-l/ contrast categorically.

On the other hand we have the findings of Eimas (1975b), who proved that two- and three month old infants from an English speaking environment (but themselves still being without active linguistic experience) are able to discriminate the /r-l/ contrast. The studies by Eimas and Miyawaki suggest respectively both the experience of an innate ability to perceive speech cues and a learning process under the influence of linguistic experience, on the basis of which perception of a certain cue can either become relevant or decline.

Some support for independence from the linguistic environment in the processing of speech in the period shortly after birth can be found in a study by Streeter (1976). She found that the English phoneme contrast between voiced and voiceless bilabial stop consonants (a contrast absent in the African Kikuyu tribe because its members only produce a prevoiced /b/) can be perceived by two month old infants from this linguistic group. Streeter used the same habituation technique and stimuli in her investigation as Eimas (1971). Consequently it is quite possible that certain speech cues are perceived at an earlier age and not later on because of the effects of linguistic experience.

It is also the very recent study by Eilers et al. (1979) that can be said to contribute to the increasing evidence that argues against the extreme assumptions of the innateness theory and that supports a more eclectic model in which experience in the learning of speech discriminations plays an important part. On the basis of their comparative study, using a habituation technique and VOT stimuli, they found that six month old infants are not always able to discriminate apriori non linguistic VOT contrasts, while adults were able to distinguish between the stimuli on the VOT continuum in accordance with their degree of contact with foreign language contrasts. Thus it appeared that adults who belonged to the class of 'phonetically sophisticated listeners' due to the fact that they were able to make narrow transcriptions of foreign languages by virtue of their profession, were capable of making far more VOT discriminations than naive listeners.

Another view of the learning of speech sound perception comes from Morse (1974). He maintains that the child, as it develops, comes to have at its disposal more and increasingly complex coding rules in its speech, ranging from low level auditory perception to the perception of very highly encoded stop consonants. For this to happen, experience with the various relevant categories is necessary.

Even Eimas (1974, 1975a) suggests the necessity for minimal experience by assuming that the innate feature detector mechanisms have to be tuned in by listening to speech, in other words have to be activated by early experience from the linguistic environment.

The theoretical arguments for assuming that speech perception in the development of the child is dependent on a learning process, are supported on more practical grounds by recent investigations into the perception of relevant speech cues in various developmental stages.

Zlatin (1974) and Zlating and Koenigsknecht (1975, 1976) investigated VOT perception and production in meaningful syllables for subjects ranging from two year old children to adults. They established that VOT perception is dependent on age, since the length of the VOT necessary to bring about the voicing contrast between prevocalic stop consonants, declines as a function of age. What is more it is not only the parameters of VOT perception which appear to be dependent on age but also production parameters. A tendency towards perceptual-motor development could be shown that revealed an increasing correspondence between perceptive and productive VOT values (see also Kent, 1976).

Similar findings are reported by Simon and Fourcin (1978) and Fourcin (1978). On the basis of their investigation into speech cue processing by English and French children from the age of two they arrive at the conclusion that categorical speech perception is not innate but results from a relatively slow learning process and, what is more, that this process has a specific pattern within each specific language area. In another investigation with children aged between 4 and 7, Simon (1975) manipulated various relevant speech cues within the speech signal. As to perception the investigation showed that younger children are guided more by the temporal organization of speech cues than by the frequency characteristics of the cues themselves. The temporal separation of categories must be greater in their case in order to discriminate satisfactorily, as if they are demonstrating a redundancy need with regard to the speech cues in comparison with older children. The acoustic frequency cues become

increasingly significant in speech perception between the ages of 4 and 7.
Eilers et al. (1975, 1977) found a developmental tendency in the perception of speech cues in children aged between 1 and 14 months. They used a visual reinforced speech discrimination paradigm and normally spoken syllables which were checked for cues via an investigation of the spectrum. Among the various findings it appeared that segments which differed by two features were discriminated earlier than segments that differed by one. These latter facts seem sufficient argument for one to accept that the categorical perception of speech is dependent jointly on experience and learning.
The probable acquisition of categorical perception of certain non-speech sounds is suggested by experiments with musical patterns offered to musicians and non-musicians. It appears that certain patterns are more clearly perceived categorically by musicians (Locke & Kellar, 1973; Burns & Ward, 1974). The dilemma discussed earlier for speech perception also applies here, that is to say, is this effect in musicians caused by talent or experience?

We have assumed in this research that categorical perception can be influenced by experience even if there are severe auditory disabilities which hamper perception. We also expect to be able to bring about an improvement in deviant perception. Examples of attempts to achieve fundamental changes in categorical perception are very scarce.
In a case study Lane and Moore (1962) describe the application of a simple operant discrimination training using extreme stimuli from a VOT continuum in the case of an aphasic patient who was no longer able to perceive voice contrast. This training was successful, as appeared from an improvement in the identification and discrimination curves of the stimuli from the continuum after training. This showed that the information obtained from the extreme stimuli can be carried over to the other stimuli within the contrasting categories in spite of the differences in VOT, and that, in accordance with one of the assumptions of the categorical perception, stimuli from the same category are not discriminated.
For discrimination learning in the case of children with profound hearing disabilities Bennet and Ling (1977) used normally spoken syllables with a voicing contrast which, on the basis of a spectographic investigation, appeared to show a clear VOT separation between related pairs. This training brought about an auditory improvement in view of the fact that after it the children showed greater capability in learning to perceive both discrete categories than before. What is more, after a thorough spectographic analysis of the VOT boundaries produced by the trained children, it appeared that one of the side effects of the training was a production improvement. Therefore it seems possible that perception training can have a spontaneous effect on production.
Carney et al. (1977) tried to influence categorical perception for subjects in a laboratory investigation. Discrimination training by means of VOT stimuli resulted in an improvement in the discrimination and identification, while by means of

selective training stimuli discrimination within categories and a phoneme boundary shift could also be achieved.

Although there are certain grounds for the assumption that deficient speech perception can be improved by means of manipulated experience, it is not clear how this learning process is realized. In this study we are not concerned with solving the dilemma as to whether the results achieved relate to the underlying neural auditory system or to the positive change of an undesirable hearing behaviour that is acquired. Our aim would be fulfilled if it could be shown that the results of the learning process are fundamental, i.e. linguistically relevant. However, in relation to developmental aspects in speech perception, Aslin and Pisoni (1978) suggest a possible explanation for some dynamic learning processes in the present research. According to their theory it is possible that during the period of development that is important for learning to perceive a certain speech feature, a so-called 'attention deficit', which exists with regard to the feature since it does not belong to the linguistic repertory of the specific language, can lead to a failure to learn to perceive the feature concerned, while the same feature might be perceptually and linguistically relevant elsewhere. The failure to actively take into consideration an attentional or productive system while learning a specific phonetic contrast can later lead to an impairment of perceptual or productive achievement with regard to this linguistic contrast. This would mean that the difficulties adults have in the discrimination of linguistic features that are not relevant for them would not be the result of a neural process as such but rather of an attentional deficit through which they have learned to ignore certain distinctive differences. In this respect we may again refer to the findings relating to the /r-l/ contrast by Eimas (1975b) and Miyawaki (1975a,b) in infants and Japanese adults and to those of Eilers et al. (1978). A learning process of this kind, however, does not lead to a definitive perceptual state. Aslin and Pisoni have shown that adults are able to categorize what are for them irrelevant speech dimensions after a short period of training. It thus also appears possible to bring about fundamental changes in speech perception in later years.
The theory of the attention deficit ties up closely with the assumptions of educationalists concerned with the auditorily handicapped child who defend the aural approach (Griffiths, 1967; Pollack, 1970; Löwe, 1970; Whetnall & Fry, 1971; van Uden, 1974; Beebe, 1978). They are unanimous in claiming that the neglect of early auditory stimulation of the residual hearing of a young auditorily handicapped child usually leads to the undesirable extinction of using the potentially usable auditory remnants, while the child comes increasingly to rely on other, principally visual, information from its surroundings. This means that auditory information becomes decreasingly relevant in favour of information deriving from other sense modalities which, incidentally, cannot compensate for deficient speech perception efficiently (see also Mark & Hardy, 1958).

The attentional state described here does not limit itself, however, in the case of a young auditorily handicapped child to the linguistically irrelevant features of the speech in its environment but concerns many relevant dimensions as a result of transmission and processing problems related to auditory patterns.

The purpose of auditory training is to make relevant or make relevant once again important information from the acoustic speech signal in precisely these cases so that this can be useful in an aural/oral communication situation. Therefore training should be started as early as possible in order to prevent extinction of adequate hearing reactions (in the sense of an attention deficit) and to optimally expose the young child with deficient hearing to speech (Fry, 1966, 1975).

Aslin and Pisoni (1978) describe four alternative theories of learning to perceive speech in which the attention deficit plays a relatively important role. No one theory has precedence over another: sometimes they can jointly explain perceptual development.

The 'Universal Theory' assumes that children are able to discriminate from birth all the possible phonetic contrasts which can occur in the speech of every natural language. Early experience then helps towards the continuation of discrimination of phonologically relevant distinctions, namely those which are actually used in the child's environment. Phonologically irrelevant contrasts do not appear in the linguistic environment, which results in the selective loss of the capacity to discriminate these contrasts. The mechanisms which are responsible for this loss of sensitivity can be either neural or attentional by nature.

The starting point of the 'Attunement Theory' is that all children can discriminate from birth at least a few of the potential phonetic contrasts in every language but that the discriminatory capacity is still incompletely developed and is probably very roughly 'tuned'. Early experience thus serves to 'tune' or to sharpen further this partially developed discriminatory capacity. Phonologically relevant contrasts from the linguistic environment are more finely tuned by linguistic experience and phonologically irrelevant contrasts remain roughly tuned or gradually weaken through the lack of specific stimulation from the environment.

The 'Perceptual Learning Theory' supposes that the capacity to discriminate whatever phonetic contrast is dependent on early experience with that contrast in the linguistic environment. The time of the development can vary depending on, inter alia, the frequency with which the phonetic contrasts occur in the early learning stage, the degree of acoustic discriminability in comparison to other contrasts and the fact of whether the child is going through a positive or a negative attentional condition.

Finally, the 'Maturational Theory' assumes that the capacity to perceive a certain phonetic contrast does not depend at all on any specific early experience but simply unfolds according to a predetermined development sequence. In this case the age at which specific phonetic contrasts can be discriminated would depend on the level of development of the underlying sensory mechanism.

The same authors also make suggestions about the different ways in which experience can selectively bring about a perceptual change. These suggestions are important for the present study since they offer a concrete opportunity to determine exactly the degree of improvement in the perception of a speech cue.

In the first place stimuli that in fact belong to two discrete perceptual categories can become more discriminable or less discriminable. These processes are called respectively *enhancement* and *attenuation*. In categorical perception they can be represented by an increase or a reduction in the peak of the discrimination curve or by an identification curve that indicates an absolute or an indistinct division into two discrete categories.

The perception of stimuli in the area of a perceptual boundary can also be tuned more finely or more crudely, which processes are respectively called *sharpening* and *broadening*. Sharpening and broadening respectively can be represented by a narrowing or a widening of the base of the discrimination peak. A narrowing of the base indicates a more abrupt perceptual transition of a phoneme to its contrasting cognate, corresponding to a relatively small transitional area on the continuum between the categories. Consequently, sharpening shows a steeper slope of the identification function between the two distinct categories, corresponding with the boundary area on the continuum. In the case of a widening of the base of the discrimination peak the perceptual transition to the neighboring category is less abrupt, the transition area on the stimulus continuum is larger and the identification curve in the transition area between the two categories shows a relatively flatter slope.

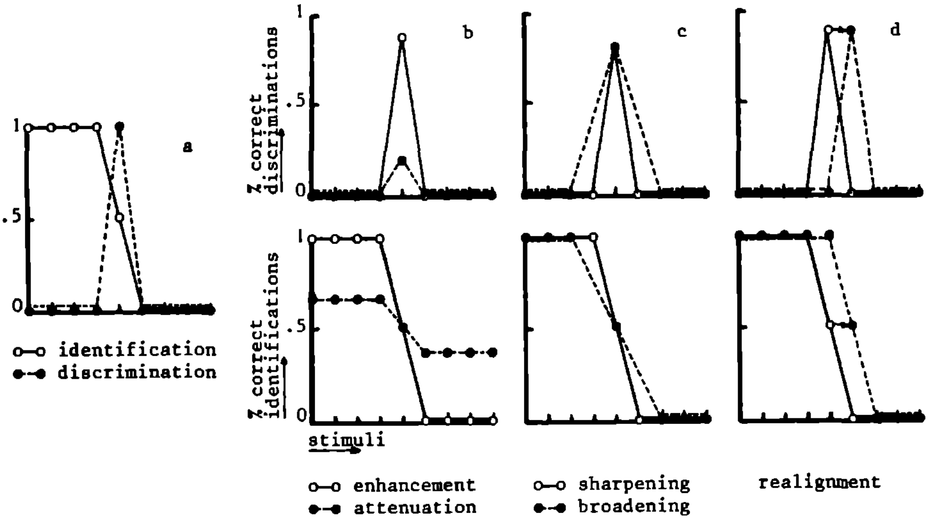

Figure 2: a) Ideal representation of an identification curve and a two-step discrimination curve on a same-different task.
b,c,d) Represent the discrimination curves (top) and identification curves (bottom) which correspond to the perceptual processes: enhancement, attenuation, sharpening, broadening and realignment.

Finally, the perceptual boundary can undergo a shift, which process is called *realignment*. This effect can be observed by means of a shift along the continuum of the peak of the discrimination curve and of the boundary area belonging to the identification curve.

In figure 2a we have a simulated ideal identification curve of a categorically perceived series of stimuli and a corrsponding two-step discrimination curve based on a two-alternative-forced-choice task. In addition to the two-step discrimination curves (top) and the identification curves (bottom) figures 2 b, c, d also display the processes of enhancement and attenuation (2b), sharpening and broadening (2c) and realignment (2d).

The terms enhancement, attenuation, sharpening, broadening and realignment will be used in this investigation in the above sense whenever an attempt is made to bring about changes in the perceptual functions of a subject which relate to them.

Summary.

Recent speech perception research in infants has lead some researchers to believe that children are born with the same capacities to perceive speech as adults and that they do this in a linguistically relevant way.

Nevertheless it is generally postulated that speech perception is influenced at an early stage by linguistic experience from the specific language environment of the child and is thus dependent on learning. The reality of the influence of learning is made acceptable by very recent findings concerning developmental phenomena which occur in the perception and production of relevant speech cues in children and by perceptual differences between children and adults.

What is more, it also appears possible so to influence perception in adults in an experimental situation that perceptually or linguistically non-relevant cues can be made relevant by manipulated experience.

The Attention Deficit Theory offers a plausible explanation for whether given speech cues become perceptually and linguistically relevant or not as a result of selective experience or the lack of it during perceptual development. The process of learning not to perceive existing differences is highly similar to the extinction phenomenon which can be observed in auditorily handicapped children who are not stimulated early enough.

Four possible processes haven been described as explanatory bases for why we perceive speech in categories: the Universal Theory, the Attunement Theory, the Perceptual Learning Theory and the Maturational Theory.

Related to these there are five different dynamic aspects of the learning of speech perception which are dealt with and which make it possible to concretize a change in learning to perceive a speech cue: the phenomena of enhancement, attenuation, sharpening, broadening and realignment. These aspects and the corresponding terms will be used in the present investigation.

2.4. HYPOTHESES.

The principal conclusions from 2.2. and 2.3. are as follows:

1. There is considerable evidence that perceptual discrimination of the voicing contrast in initial stop consonants is based on the extraction of the distinctive characteristics [+voicing] and [-voicing] respectively.
2. The perception of speech cues can be influenced by specific experience, irrespective of age.
3. The influencing of the perception of speech cues through manipulated experience can be established objectively by means of the categorical perception paradigm.

These claims are seen as the basis of the following hypotheses:

Hypothesis 1: Deviation in one or more parameters of the phoneme boundary area belonging to the VOT continuum that lies at the base of the voicing contrast of word-initial prevocalic stop consonants in the alveolar pair /dɑk-tɑk/ or the bilabial pair /bɑk-pɑk/, can be improved in many cases by specific auditory discrimination training. Such improvement may be based on perception modifications which are the result of enhancement, attenuation, sharpening and realignment brought about by training.

The four reference parameters which belong to the phoneme boundary area of the /d-t/ or /b-p/ VOT continuum and which are being used as the norm, are the upper limit, the lower limit, the cross-over point and the phoneme boundary zone. They are established according to criteria described by Zlatin and Koenigsknecht (1975). They consist of linear interpolations of boundary values on the VOT continuum which are associated with certain response values on the identification curves obtained from subjects who were not auditorily handicapped and who perceived the relevant stimuli in normal categorical fashion.

On the basis of the stimuli identified as voiced and the corresponding identification curve, the phoneme boundary parameters can be described as follows:

a. The lower limit is the actual point or the point obtained through linear interpolation on the VOT continuum that corresponds to 75% of the total number of presentations where a stimulus from this continuum is identified as voiced.

b. The upper limit is the actual point or the point obtained through linear interpolation on the VOT continuum that corresponds to 25% of the total number of presentations where a stimulus from this continuum is identified as voiced.

c. The cross-over point on the VOT continuum is associated with the stimulus or

stimulus value obtained through linear interpolation where 50% of the number of presentations is perceived as voiced and 50% is perceived as voiceless.
(The points in a, b and c are given in msec).

d. The phoneme boundary zone is established by subtracting the lower limit from the upper limit and is also given in msec.

e. The stimuli where 75% or more of the total number of presentations are identified as voiced are counted as belonging to the voiced category in the learning experiment, provided that at least two neighboring stimuli obtain this percentage or more. The stimuli where 25% or less of the total number of presentations were identified as voiced were counted as belonging to the voiceless category in the learning experiment, provided that at least two neighboring stimuli obtained this percentage or less. Figure 3 illustrates the phoneme boundary parameters of a simulated curve based on VOT stimuli that have been perceived as voiced.

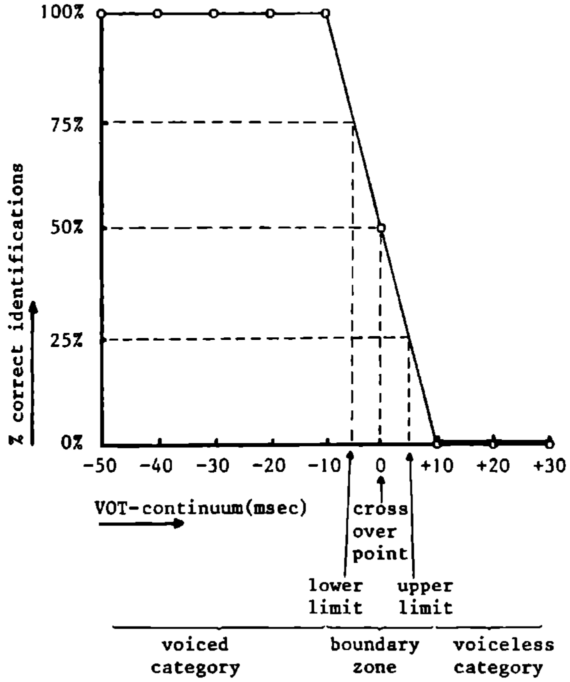

Figure 3: Phoneme boundary parameters of a simulated VOT continuum and the corresponding [+voice] identification curve.

Hypothesis 2: If the categorical perception of the voicing contrast in the above mentioned word-initial, prevocalic stop consonants is improved by means of a special discrimination training using controlled stimuli, this perceptual experience also brings about a broader, linguistically relevant effect. This effect primarily concerns the transfer of the perceptual improvement to normally spoken prevocalic word-initial stop consonants that were not part of the training and that are produced in the same place of articulation but occur in a different vowel context. Furthermore it is considered linguistically relevant if a generalization occurs from the improved perception of the voicing features in the above stop consonants to normally spoken prevocalic word-initial stop consonants that were not part of the training and that differ both as to the place of articulation and the vowel context. The underlying assumption of hypothesis 2 is that an improved extraction, resulting from auditory training,of the feature 'voiced' or 'voiceless' from a specific speech signal is not limited to initial consonants that share this feature and are produced in the same place of articulation as the signal, but that it also includes initial consonants with these features that are produced in a different place of articulation. What is more, a certain context independence of the voicing features is given extra emphasis by predicting transfer of information to the same stop consonants in various vowel contexts.

In contrast to hypothesis 1 in which the dependent experimental variable consists of the phoneme boundary parameters of a VOT continuum, the dependent variable in hypothesis 2 is determined by the number of [+voice] and [-voice] features occurring in word-initial prevocalic stop consonants which are part of normally spoken minimal pairs with the voicing feature as critical variable, and which are included in a /d-t/ or /b-p/ list.

A third list used for the analysis of the improvement of the voicing feature extraction contains a number of [voice x place] contrasts /b-p-d-t/.

CHAPTER 3

CONSTRUCTION OF THE STIMULI AND DETERMINATION OF THE PHONEME BOUNDARIES

3.1. THE STIMULI.

In their study of the voicing dimension in various languages Lisker and Abramson (1964) found that the VOT for realizing an initial voiced consonant often precedes the burst release and the VOT of a voiceless initial consonant follows for a two-category language, that is, a language in which the manner in which a stop consonant is produced can be divided into voiced and voiceless categories. The perception of voicing contrast was closely correlated with this: voiced judgements of stimuli were associated with a voice lead and voiceless judgements with a voice lag. In Dutch initial stop consonants, the perceptual transition from the voiced to the voiceless category, expressed in VOT, occurs abruptly and very shortly after the burst portion, viz. between 0 and 25 msec.

Slis and Cohen (1969a) found that, even with synthetically constructed speech stimuli, the judgement of voiced stimuli correlated with a voice lead. Dutch listeners made the best judgement of the voiced characteristic when the value of the voice lead was -80 msec VOT. The most accurate judgement of a voiceless stimulus was at a value of +10 msec voice lag.

In general the phoneme boundary area between voiced and voiceless categories shows a very abrupt transition. Among selected, naive listeners, Liberman et al. (1958, 1961) reported a transition from a voiced judgement to a voiceless judgement of within 10 msec when a cut-back method of the first formant was applied to the stimuli.

In their attempt to determine some acoustic dimensions that are relevant for perception of the voicing distinction in consonants, Slis and Cohen (1969b) made use of synthetically produced consonant-vowel stimuli,the consonants of which belonged to the labial and alveolar category. They found that the consonant part of a complex stimulus which was composed of voice lead, burst release, formant transitions and vowel part was perceived 100% of the time as voiced. The voice lead seemed to be very relevant for the perception of the voiced characteristic of such stimuli. But the consonant part of the same stimulus without voice lead and with a by segmentation produced silent-gap after the burst release was perceived 100% of the time as voiceless. Here it seemed that the silent-gap had perceptual importance.

These results formed our basis for the construction of two VOT-series. The underlying thought was that is should be possible by way of a segmentation procedure and using an original voiced stop consonant, to realize a perceptual transition from the voiced category to the voiceless category. A segmentation procedure seemed to be a useful and relatively simple way to construct a VOT continuum with the goal of determining the correlation between acoustical and perceptual cues, using subjective measurements ('t Hart & Cohen, 1964; Nooteboom & Cohen, 1976).

This was actually not the only reason for choosing this procedure. Another assumption was that stimuli composed of actual speech, despite alterations, would also sound more speech-like than would the relatively unnatural sounding stimuli which were strictly synthesized by rule. This assumption was considered important for the application of manipulated stimuli in a group of auditorily handicapped children.

Finally, it seemed also necessary to construct Dutch stimuli, since the phoneme boundary values and the relevant speech cues of the frequently used synthetically produced English VOT continua do not correspond with the Dutch language (Lisker & Abramson, 1964; Lisker, 1970). Application of an English VOT continuum, in which the phoneme shift is perceived relatively late after the burst release, has been shown to lead to an undesireable artificial broadening effect in the perceptual functions among listeners from a language area where this shift follows immediately after the burst (Basso et al., 1977).

Next to the use of actual speech in the construction of the stimuli, care was taken that the critical consonants /d-t/ and /b-p/ functioned as initial phonemes of two minimal pairs, consisting of four meaningful words: [bɑk] <bin> , [pɑk] <parcel> , [dɑk] <roof> and [tɑk] <branch> .

By choosing meaningful words, it was later possible to pose the multiple choice questions in the form of pictures for children who could not yet read.

A PDP 11/45 computer-controlled gating technique was used for the stimulus construction by which it was possible to isolate segments of speech uterances of any desired length.

Two isolated spoken CVC[*)] syllables,produced by a male voice, functioned as the basic stimuli. One began with a voiced bilabial consonant in [bɑk] and the other began with a voiced alveolar consonant in [dɑk]. These syllables were recorded with a Revox A-77 tape recorder.

Both speech signals were fed through a band filter of which the lower limit was set at 50 Hz and the upper limit at 5000 Hz. Then they were digitalized via an analog-digital converter and stored in the computer. The sample frequency was 1666 Hz. In this way, the entire waveform or parts of the speech signal could be recalled at any desired moment. Retrieval was possible in two fashions: an auditory representation via a loud-speaker or a visual display of all or part of the waveform on a monitor. This procedure made it possible to manipulate the stored stimuli as desired, and to edit the two VOT series. When a particular processed stimulus satisfied the established requirements, the signal was rerecorded on audio tape via a digital-analog converter and the 50 - 5000 Hz filter, in order to construct label and discrimination lists.

The following steps were performed sequentially. First, the length and the position of the burst release were determined for both initial stimuli [bɑk] and [dɑk]. Because the same burst portion had to serve for the voiced and unvoiced cognate

[*)] CVC = consonant-vowel-consonant.

stimuli in the VOT series to be constructed, and because it is usually longer in a voiceless consonant than in a voiced consonant, a neutral middle value was chosen for the burst portion. In the study of Slis and Cohen (1969a) synthetically produced stimuli with a fixed burst portion of 10 msec were used to produce a noticeable voicing contrast.They reported also a clearly perceivable voicing contrast if a burst of 5 msec was used for a voiced stop consonant in a consonant-vowel context and a burst of 20 msec for a voiceless cognate consonant (1969b). In our study, a standard burst portion of 12 msec was selected from the word [bɑk] for the construction of voiced and voiceless stimuli of the VOT series [bɑk-pɑk] with the use of the visually displayed waveform. A burst portion of the same length was selected from the word [dɑk] for the construction of the VOT series [dɑk-tɑk]. The burst portions for both series were used as invariable cues in order to allow the speech stimuli to vary only on the basis of a variable VOT cue.

After determining the length and the position of the two burst portions, the voice lead, which immediately preceded the onset of the burst release, was reduced in intervals of 10 msec by means of a successive segmentation procedure. For the /b-p/ continuum, a maximum prevoicing time of 50 msec was chosen. A maximum prevoicing time of 30 msec was chosen for the /d-t/ continuum. No voice leads longer than 50 respectively 30 msec were used, contrary to what was reported by Lisker and Abramson (1964), Lisker (1970) and Slis and Cohen (1969a).

The most important motivation for the choice of a shorter prevoicing period was that the length of the identification and discrimination series, as a result of the stimulus reduction, could be considerably shortened, so that the assignments of the series was less lengthy and more accessable for severely auditorily handicapped children. By this stimulus reduction, we tried to obtain two contrasting categories in the identification test and the desired trough and peak effect in the discrimination test using a minimum number of stimuli. Furthermore, a preliminary listening test with normal listeners showed that they were able to assign the quality of 'voiced' to stimuli of the /d-t/ series within the range of a prevoicing period of -30 to 0 msec.In the /b-p/ continuum it turned out that a prevoicing period of -50 to 0 msec was required.

After determining the burst and prevoicing period the total duration of the first stimuli from the two series was 275 msec for [bɑk] and 320 msec for [dɑk].

The successive reduction of the voice lead in intervals of 10 msec VOT produced six stimuli for the /b-p/ series in the -50 to 0 msec range on the VOT continuum situated relative to the onset of the burst release. There were four stimuli in the -30 to 0 msec range in the /d-t/ series.

In order to obtain a voice lag effect, a silent interval was substituted for the vowel, immediately after the 12 msec-long burst release within the words [bɑk] and [dɑk]. This interval varied from 10 to 30 msec in intervals of 10 msec. Consequently the effective voice lag lies in a range from 12 to 42 msec on the continuum.

In total, the /b-p/ VOT series consisted of a set of 9 different VOT stimuli (-50, -40, -30, -20, -10, 0 msec voice lead and +10, +20, +30 msec voice lag after the

12 msec burst period). The /d-t/ VOT series consisted of a set of 7 different VOT stimuli (-30, -20, -10, 0 msec voice lead and +10, +20, +30 msec voice lag after the 12 msec period of the burst release).

In actuality, the critical time cues in both series are not exactly the same as in VOT continua which are synthesized according to rule and where the differences between the stimuli can be kept entirely constant.

Not all of the VOT intervals are 10 msec long in our series. In the /b-p/ series there exists an equal VOT interval of 10 msec between the first six and the last three stimuli. In the /d-t/ series there is a 10 msec interval between the first four and the last three stimuli.

The interval between stimulus 6 and 7 in the /b-p/ series and stimulus 4 and 5 in the /d-t/ series consists of 12 msec burst portion and 10 msec voice lag, a total of 22 msec, and covers a range of 0 to 22 msec of the VOT continuum. These longer intervals are unavoidable with the segmentation procedure used, in view of the fact that the burst release is indispensable for the perceptual judgement of the stimulus as an initial stop consonant, and may not be removed. A total or partial elimination of the initial burst release from the CVC speech signal would produce a sound at the beginning of the vowel that approximates a hardly identifiable glottal-stop (see Pols & Schouten, 1978).

The disadvantage of an irregularity in one of the intervals of both VOT series concerns especially the application of the categorical perception paradigm.

With a labeling procedure, in which every isolated stimulus presented must be judged as 'voiced' or 'voiceless', the phoneme boundary values will fall more or less differently than in a continuum with equal intervals, because of the irregularity that appears in the phoneme boundary region. The appearance of the perceptual phoneme shift from the judgements 'voiced' to 'voiceless' is expected for both series in the VOT range immediately after the burst period. As a result of an always present burst portion of 12 msec, this shift always corresponds with a larger phoneme boundary area than when the VOT intervals had been exactly 10 msec after the burst onset. This artificial and relative small broadening effect of the phoneme boundary area applies for all normal listeners and therefore is not considered a draw-back. This irregularity is taken into account in the interpolation of boundary parameters on the basis of perceptual functions.

The unequal interval concerns also the discrimination procedure in which judgements are made whether two successive stimulus words are perceived as the same or different words. The time differences of the one, two and three step comparisons were 10, 20 and 30 msec VOT respectively. The differences of the comparisons over the phoneme boundaries are however greater, that is 22 msec in the one-step comparison, 32 msec in the two-step comparison, and 42 msec in the three-step comparison. For the two-step comparisons these differences were 32 msec between stimulus 3 and 5 in the /d-t/ series and stimulus 5 and 7 in the /b-p/ series (corresponding with -10 and +22 msec VOT), and between stimulus 4 and 6 in the /d-t/ series and stimulus 6 and 8 in the /b-p/ series (corresponding with 0 and 32 msec VOT). The two-step comparisons will be used in this study. In spite of the irregularity between the intervals on the VOT continuum, the two-step-intra-category comparisons are based on equal VOT intervals. Also the two-step-inter-category comparisons over the phoneme boundary are based on equal but slightly longer intervals. In this investigation, the appearance of reliable categories, peaks and troughs in the identification and discrimination functions will be considered as most important.

In summary, we can say that we expected that, despite the presence of an unequal interval in the two series, the labeling results would show a desireable phoneme switch within the normal phoneme boundary area but with a slightly larger separation between the two categories. Likewise, we expected that the discrimination results would produce the peaks and troughs that would correspond with the observed identifications.

The two sets of stimuli were recorded on DEC tape for the automatic construction of randomized label and discrimination lists.

The waveforms of the words [bɑk] and [dɑk] are shown in figure $4a_1$ and $4a_2$. The intersecting lines correspond with the successively performed segmentations. The voice lead can be seen preceding the burst portions, while a total voice lag of maximally 30 msec is substituted for the vowel portion after the burst portions.

In figure $4b,_{1-4}$, are the spectographic representations of extreme stimuli of the /b-p/ series, respectively -50 msec voice lead and +30 msec voice lag and those of the /d-t/ series, respectively -30 msec voice lead and +30 msec voice lag.

3.2. CONSTRUCTION OF THE PHONEME LABEL AND DISCRIMINATION LISTS.

With the help of a PDP 11/45 computer-controlled randomizing program, the nine stimuli from the /b-p/ series and the seven stimuli from the /d-t/ series were recorded in random order on tape by a Revox A-77 recorder in such a way that separate phoneme label and discrimination lists were obtained.

3.2.1. The /b-p/ stimuli.

a) Out of the nine /b-p/ stimuli, five randomized blocks were recorded, each consisting of 18 stimuli. Every stimulus appeared twice in each block. The interstimulus interval was 4 sec. Preceding the five blocks, extreme stimuli 1 and 9, having a VOT-value of -50 and +42 msec respectively, were each recorded eight times in succession with interstimulus intervals of 3 sec. These were used as example stimuli. Before each block the number of the block was mentioned.

b) For the discrimination lists, preference was given to a two-alternative-forced-choice task consisting of same-different judgements, as opposed to the ABX-procedure which usually is applied in categorical perception experiments.

The same-different judgement is a less complex task and is less dependent on the auditory short memory. It has no disadvantage compared with a ABX-judgement such that

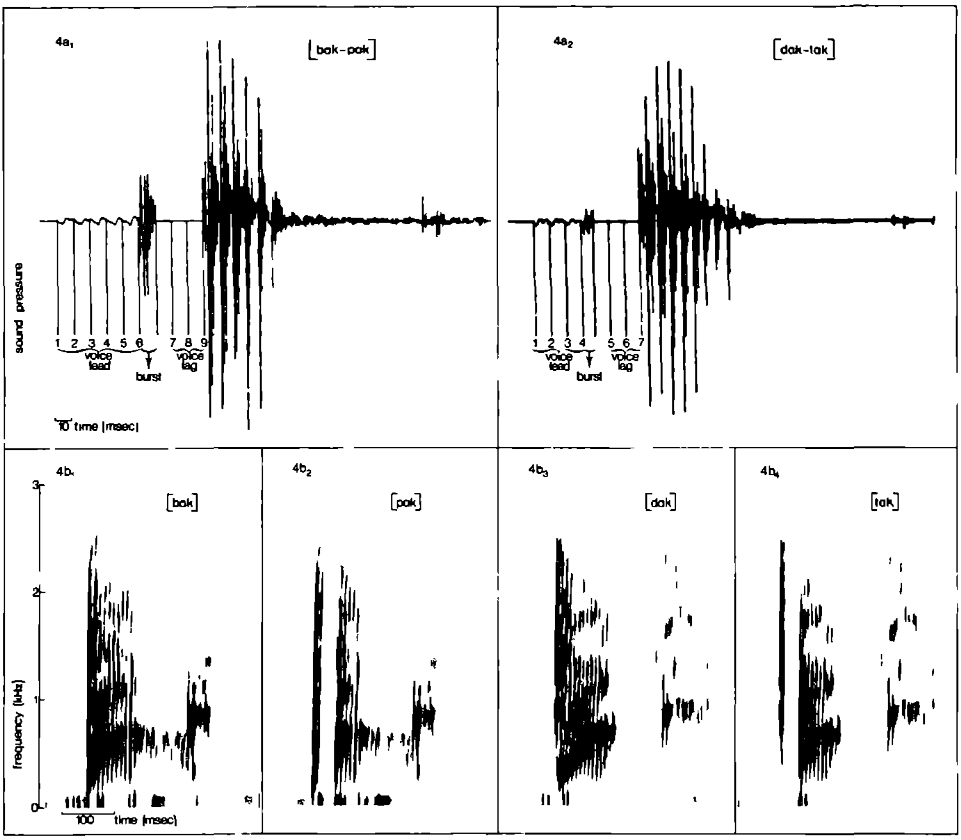

Figure 4: a) Wave forms of the words [bɑk] and [dɑk] with the intersecting lines of the performed segmentations. Left of the burst portion the voice lead can be seen and to the right of it the maximal silent gap.

b) Spectograms of extreme stimuli from the /b-p/ and /d-t/ series (Sound Spectrograph: Voice Identification, series 700).

(b1) belongs to stimulus 1, [bɑk] , with -50 msec voice lead; (b2) to stimulus 9, [pɑk] , with +30 msec voice lag after the 12 msec burst portion; (b3) to stimulus 1, [dɑk] , with -30 msec voice lead; (b4) to stimulus 7, [tɑk] , with +30 msec voice lag after the 12 msec burst portion.

it lends itself better to use with normal children and subjects with brain dysfunction or mental retardation (Pollack & Pisoni, 1971; Zlatin & Koenigsknecht, 1975; Basso et al., 1977; Blumstein et al., 1977).
Each of the possible one-, two- and three-step comparisons between the stimuli of the VOT series was included in the discrimination lists for the normal subjects. For use with the auditorily handicapped subjects, a separate list was constructed with two-step comparisons only.
The nine stimuli from the /b-p/ VOT series made it possible to construct eight different one-step, seven different two-step and six different three-step comparisons, a total of 21 comparisons.
Of these 21 comparisons, eight random blocks were recorded such that each comparison occurred once per block. Care was taken to record each comparison four times in the AX- and four times in the XA-order, to eliminate response asymmetry (cf. Uselding, 1978). The time interval within an AX-comparison was set at 800 msec. The interval between two comparisons (the response time) was 4500 msec. Before each block, the number of the block was mentioned.

c) Four separate random blocks of fourteen two-step comparisons per block were recorded on tape for the auditorily handicapped subjects. In a block each two-step comparison occurred once in the AX- and once in the XA-order. Intra- and inter-comparison intervals were also set at 800 and 4500 msec respectively. Before each block the number of the block was mentioned.

3.2.2. The /d-t/ stimuli.

a) For the construction of random label lists of the /d-t/ VOT series, the same procedure was followed and the same times were used as for the /b-p/ series, with the exception that the random blocks now consisted of seven stimuli instead of nine. The five randomized label blocks consisted of 14 stimuli each. In each list a certain stimulus occurred twice. The extreme stimuli 1 and 7, having a VOT value of -30 and +42 msec respectively, were each recorded eight times in succession with interstimulus intervals of 3 seconds and used as example stimuli preceding the five random blocks.

b) Six one-step, five two-step and four three-step comparisons can be made with 7 stimuli, a total of 15 comparisons. Eight random blocks, each consisting of 15 comparisons, were recorded. Each comparison occurred once per block. Each comparison occurred a total of four times in the AX- and four times in the XA-order.

c) Four separate random blocks, each block consisting of ten two-step comparisons, were recorded for the auditorily handicapped subjects. Every two-step comparison occurred once in the AX- and once in the XA-order per block. Intra- and inter-comparison intervals were 800 and 4500 msec respectively. Before each block the number of the block was mentioned.

3.3. DETERMINATION OF THE /D-T/ AND /B-P/ PHONEME BOUNDARY VALUES.

3.3.1. Data collection.

3.3.1.1. The subjects.

As subjects for obtaining normal values for the phoneme boundaries, 33 listeners ages ranging from 4 to 41 years and having a non-dialectic pronounciation of the Dutch language, were used. Zlatin and Koenigsknecht (1975) found no age effects in their study of perceptual VOT phoneme boundary values of labial and dental, prevocal stop consonants among American adults and 6 year old children. Perceptual differences were found between 2 year old children and listeners who were 6 years old or older. In our study, the lower limit for age was set at 4 years for the normal listeners. Due to a lack of information about the perception of temporal speech cues at an older age, no upper limit was given to the age requirement.

The initial group of normal listeners consisted of 6 boys varying in ages from 4 to 9 years old, 7 girls ages 5 to 10 years old, 8 men ages 22 to 41 years old and 12 women ages 19 to 34 years old. These subjects were not familiar with specific listening tests. They had unimpaired hearing and speech and normal intelligence, as could be estimated on the basis of their school records, education, and occupations.

Two children, a four year old boy and a five year old girl could not yet read or write and two girls and boys were beginning readers. A special strategy was used for these non-readers and beginning-readers, which deviated from the normal procedure for the listening experiment. Participation in the experiment was voluntary. No material rewards were given.

3.3.1.2. Materials.

The auditory stimulus material is described in section 3.2. Various lists were designed, corresponding with the auditory stimulus series, which were based on the two-alternative-forced-choice principle.

For labeling the /b-p/ phonemes, a worksheet was used on which 5 blocks of 18 choice pairs <bak-pak> each were printed. These blocks coincided with the auditory stimulus series numbers 1 through 5. A picture of a bin ([bɑk] in Dutch) and one of a parcel ([pɑk] in Dutch) were showed at the top of each worksheet along with the printed words <bak> and <pak> . These pictures and words could be referred to during the labeling task if the association between speech sound, printed letter, and concrete visual representation created problems. This was very helpful for several auditorily impaired children. For the children who could not read or could only read poorly, two separate pictures of a bin and a parcel respectively were used as concrete choice material during the labeling task, which, to prevent response bias, could be presented to the child in any desired horizontal or vertical position. In this case the subjects did not mark their choices on the sheet themselves, but their responses were recorded by the experimenter on the same worksheet used by the older listeners.

For labeling the /d-t/ phonemes, a worksheet was used with 5 numbered blocks of 14 choice pairs <dak-tak>. At the top of this worksheet were two pictures, one of a roof ([dɑk] in Dutch) and one of a branch ([tɑk] in Dutch) along with the corresponding printed words <dak> and <tak> as reference. The younger children received the alternative choices in the form of concrete pictures.

The /b-p/ discrimination test was conducted using a worksheet with 8 sets of 21 choice pairs with the alternative <s-d> (same-different). The 8 sets were numbered 1 through 8. For the /d-t/ discrimination test, a worksheet was used consisting of 8 sets of 15 choice pairs with the <s-d> alternative.

The examiner recorded the verbal 'same or different' responses of the non-readers and beginning-readers on the same worksheet as was used by the other subjects. Furthermore the non-readers made the choice of 'same' or 'different' using two different pictures, one consisting of two identical visual symbols (o o for 'same') and two different symbols (o ■ for 'different'). The concepts 'same' and 'different' were connected to the configurations during a short pretraining session.

A Revox A-77 tape recorder and a set of AKG K150 headphones, equipped with circumaural cushions, were used for presenting the stimuli. The stimuli were presented binaurally at a volume which was most comfortable for each subject, as was previously determined by the subject himself. A subjective adjustment of the volume of presentation, instead of a frequently used standardized sound level, was not thought to cause an important variation in response. The most comfortable level varied from 65 to 90 dB SPL. Recently it has been shown that phoneme contrasts, based on a VOT variable show no significant response variation when offered on various volume levels in a labeling task, as long as the sound level exceeds 35 dB SPL (Simon, 1978).

The examiner followed the test procedure using his own set of headphones and could intervene when necessary.

3.3.1.3. Procedure.

The listening tests were conducted individually in a quiet room. A session of approximately 40 minuts a day was spent per subject. An entire identification test of 5 blocks and an entire discrimination test of 8 blocks of the same related pair were presented in each session. It was necessary to shorten the sessions with some of the younger children. Besides the randomization of the auditory stimuli within each block, the order of presentation of the tests and of the blocks within each test was randomized. The order of presentation of the identification and discrimination tests of a related phoneme pair was randomized as well as the order of presentation of a given /b-p/ or /d-t/ test.

The listener was asked to make a quick decision and to mark his choice with a pencil, without producing a vocal imitation of the stimulus.

The first two labeling tests of the /d-t/ and /b-p/ series were preceded by eight repetitions of the extreme stimuli, spoken by the examiner. Preceding the first discrimination test the possible stimulus combinations [bɑk-pɑk] , [pɑk-bɑk] , [bɑk-bɑk] and [pɑk-pɑk] were presented orally by the examiner along with the corresponding choices 'same' <s> or 'different' <d> .

The non-readers and beginning-readers were shown how to point to the pictures used for the identification and discrimination tests. The positions of the pictures were continually changed during the tests to avoid response bias. Instruction was given in the correct choice of 'same' and 'different' symbols using the possible stimulus combinations. The tape recorder could be stopped at any moment during a session. Labeling and discrimination tests were presented four times for each listener. In this way, 40 [+voice] or [-voice] identifications for each label stimulus and 32 'same-different' judgements for each discrimination pair were obtained. Thus, for 33 listeners, the total number of identifications per stimulus was 1320 and the total number of discriminations per stimulus pair was 1056.

3.3.2. Data reduction.

The identification curve, based on the [+voice] judgements of the /b-p/ and /d-t/ series was determined for each listener in the beginning group (n= 33). Forty judgements per stimulus were obtained for each listener.
The lower limit, the 50% cross-over point, the upper limit and the phoneme boundary zone were calculated in msec for the 33 phoneme boundary areas. In this way, 8 values were obtained per listener, 4 belonging to the /b-p/ series and 4 belonging to the /d-t/ series. In connection with determining the normal values of these phoneme boundary parameters,it was uncertain whether the 33 observed values could be pooled for each parameter. Possible differences in the data among the subgroups had to be checked, in other words whether an age or sex effect was present in the data. This question was answered using an analysis of variance.
Furthermore,there was the problem of whether a VOT difference existed between the phoneme boundaries of the /b-p/ and /d-t/ series in the sense of a 'place of articulation' effect.
Lisker and Abramson (1964) were first to demonstrate the systematic movement of the productive and perceptual voicing phoneme boundaries on the VOT continuum in the direction of the burst portion when the place of articulation of two related stops changes from a back to a front position within the vocal tract. Thus we would expect that the position of the /b-p/ phoneme boundary on the VOT continuum would be nearer to the burst release than the position of the /d-t/ boundary,corresponding with a smaller voice lag. This implies a temporal variation and has consequences for a possible transfer of learned voicing information toward phonemes which differ in the place of articulation. A paired-t statistical procedure was used for this problem.
For both the analysis of variance and the paired-t test, the assumption was made that the samples were normally distributed. The relative measure of skewness for the corresponding samples was calculated for this.
To analyze the sex-age effect, the 33 calculated scores from each phoneme boundary value were divided into the following sub-groups: boys (n= 6), girls (n= 7), men (n= 8) and women (n= 12).

These sub-groups were then checked for homogeneity. For each sex-age sub-group and for each phoneme boundary value the outliers were eliminated according to Chauvenets criterion E, (Documenta Geigy, 1959). The outlier-test established the outliers for each sample on the basis of deviations from the mean in relation to the standard deviation. If outliers are included in a sample they affect the mean and standard deviation.

The final criterion for an elimination was that outlying values had to be measured for a listener for more than one parameter of a given phoneme boundary area, including the phoneme boundary zone. If this was the case, all data for such a listener were eliminated.

After elimination of the subjects with outlying values a means and moments program was used to calculate the relative measure of skewness of the distribution of all sex-age sub-groups for each of the eight phoneme boundary parameters (Spiegel, 1961).

A relative value of less than -0.5 or greater than +0.5, but lying within the limits of +1.0 and -1.0,was established eighteen times. Fourteen distributions had a relative value of less than +0.5 and greater than -0.5. Because a distribution may be considered to be symmetric when $-0.5 < \text{Skewness} < +0.5$ and highly skewed when this value exceeds ± 1.0, it can be said that the distributions varied from approximately normal to slightly skewed, while there were no exceptional skewed distributions in the sample of sub-groups.

A one-way analysis of variance was then preformed on the data of the sub-groups (sex-age) with the hypothesis that there were no significant differences between these groups. No significant differences were found except for the phoneme boundary zone of the /d-t/ series where the boundary value of the men differed from the rest. These zone values were still used for further calculations in view of the fact that there were no other systematic deviations found for the sub-group of men.

The results of the analysis of variance made it possible to combine the individual scores for each phoneme boundary value.

An outlying value test was made for all of the 33 pooled data per boundary value, in view of the fact that this test for the combined data could lead to the elimination of listeners other than those which might be eliminated by an outlying values test of the sub-groups. As was expected, a great agreement was found between the outlying values of the sub-groups and that of the total group. Nine subjects were eliminated on the basis of this outlying values test: 1 boy, 1 girl, 3 men and 4 women. The number of remaining listeners was 24. The age range of the sex-age sub-groups was not changed by the elimination process.

The relative measure of skewness, calculated from the pooled data of the 24 remaining listeners, yield for the 8 phoneme boundary areas 4 slightly skewed distributions and 4 distributions which are approximately normal, while there appear no exceptionally skewed distributions.

Tabel 1 shows the number of listeners and the mean, standard deviation and standard error of the mean in msec for all 8 phoneme boundary values for each sub-group. The results of the one-way analysis of variance are also shown.

/b-p/series					/d-t/series			
75%	50%	25%	zone		75%	50%	25%	zone
				n boys=5				
1,40	8,36	15,72	14,32	Mean	0,80	7,80	15,50	14,68
3,94	2,76	1,71	3,60	SD	5,61	5,76	4,45	2,44
1,76	1,24	0,77	1,61	SeM	2,51	2,57	1.99	1,09
				n girls=6				
2,62	8,65	15,82	15,00	Mean	4,13	11,80	19,82	15,68
4,23	4,48	6,05	2,39	SD	2,38	2,04	3,18	2,51
1,73	1,83	2,47	0,98	SeM	0,97	0,83	1,30	1,02
				n men=5				
1,88	7,90	16,14	14,28	Mean	4,52	10,74	15,80	11,30
4,08	4,56	4,43	2,86	SD	1,23	1,97	0,67	0,82
1,82	2,04	1,98	1,28	SeM	0,55	0,88	0,30	0,36
				n women=8				
1,24	8,31	16,14	14,91	Mean	4,40	10,83	19,01	14,53
5,16	6,18	4,47	3,83	SD	3,60	4,09	5,06	2,39
1,82	2,18	1,58	1,35	SeM	1,27	1,45	1,79	0,85
.12	.02	.20	.08	F(3,20)	1.34	1.11	1.77	3.89
>.10	>.10	>.10	>.10	p	>.10	>.10	>.10	.05

Table 1: Number of listeners per sub-group sex-age and the mean, standard deviation and standard error of the mean of 4 VOT phoneme boundary values of the /b-p/ and /d-t/ series.
Below are the results of the analysis of variance on the data of the sub-group sex-age for each boundary value: the F-values with 3 degrees of freedom for variation between the samples and 20 for total variance and the two-tailed levels of significance are given.

In table 2 the averages in msec with the standard deviations and the standard errors of the mean for each phoneme boundary value are shown, calculated for the pooled results of all 24 listeners. The final norms are based on these data. These have a Students-t distribution. The 95% confidence interval was determined for each boundary value according to the formula: mean ± (t.SD) where t= 2,069 when 2p= .05 and when there are 23 degrees of freedom.

In table 2 below are shown the 8 minimum-maximum confidence levels of these intervals. These are used further as reference data in the learning research in chapter 5.

/b-p/series					/d-t/series			
75%	50%	25%	zone	n=24	75%	50%	25%	zone
1,75	8,32	16,39	14,68	Mean	3,61	10,42	17,81	14,18
4,23	4,58	3,92	3,08	SD	3,62	3,81	4,14	2,60
0,86	0,94	0,80	0,63	SeM	0,74	0,78	0,84	0,53
-7,00	-1,16	+8,28	-	t=2,069 df=23 min.-max. of 95% confidence interval	-3,88	+2,54	+9,24	-
+10,50	+17,80	+24,50	21,05		+11,10	+18,30	+26,38	19,56

Table 2: *Above:* means, standard deviations and standard errors of the mean of each phoneme boundary parameter, calculated for the pooled results of the 24 listeners.
Below: the minimum and maximum limits, belonging to the 95% confidence intervals of the phoneme boundary values.

3.4. THE PERCEPTUAL FUNCTIONS.

3.4.1. The identification and discrimination curves.

In section 2.1. it is hypothesized that the /b-p/ and /d-t/ series had to be perceived in accordance with the assumptions of the categorical perception. To show this, the identification and two-step discrimination scores of the 9 /b-p/ and the 7 /d-t/ stimuli were pooled for the 24 remaining listeners.
The two-step discrimination scores were used further in this investigation because later it turned out that the auditorily handicapped subjects had too much difficulty with the one-step discriminations and because the three-step discriminations produced too little information for the desired peak and trough forming.
The [+voice] identification score was used further in this investigation. The pooled identification scores are based on 960 judgements for each stimulus. Each pooled two-step discrimination is based on 768 comparisons.
The number of [+voice] identifications for each stimulus and the number of 'different' responses for each comparison were converted to percentages of the total number of judgements. These percentages are shown for the two stimulus series in table 3.

/b-p/stimulus	1	2	3	4	5	6	7	8	9
% /b/responses	98,75	98,63	98,55	95,67	93,50	76,20	9,71	3,67	4,92
discrimination step	1-3	2-4	3-5	4-6	5-7	6-8	7-9		
% observed 'different' responses	5,91	10,91	14,45	35,09	80,64	79,10	24,18		
% predicted 'different' responses	2,27	5,00	8,64	22,91	83,55	76,10	14,27		
/d-t/stimulus	1	2	3	4	5	6	7		
% /d/responses	98,08	98,92	95,58	87,63	12,88	2,29	1,42		
discrimination step	1-3	2-4	3-5	4-6	5-7				
% observed 'different' responses	3,73	16,55	90,90	91,27	23,18				
% predicted 'different' responses	3,18	15,00	89,54	85,36	9,09				

Table 3: The percentages of the stimuli labeled as [+voice] and of the two-step comparisons judged as 'different' from the /b-p/ and /d-t/ series. Each label score is based on 960 judgements and each discrimination score is based on 768 judgements of a total of 24 listeners.

In order to obtain the predicted discrimination values it was assumed that a listener can discriminate between stimuli only to the extent to which he can identify them as different phonemes. For this, use was made of the observed responses in the phoneme labeling tasks as a basis for the calculation of the predicted frequency with which a listener can make the correct discrimination between each presented XA arrangement of the stimuli (cf. Liberman et al., 1957, 1961a; especially Pollack & Pisoni for the same-different paradigm, 1971). Thus,if each of the two stimuli of a pair was assigned the same label in the identification test, the assumption was that the listener would judge this pair as 'same' in the discrimination test. If each of the two stimuli received a different label in the identification test, it was expected that the listener would answer with 'different' in the discrimination test.

The obtained results are the average predicted percentages 'different' responses, calculated for 24 listeners. These are also shown in table 3.

The perceptual functions for the /b-p/ and /d-t/ series, corresponding with the data in table 3, are shown in figures 5(1) and 5(2). The points of the two-step discrimination curves are based upon comparisons between two VOT stimuli of the continuum (for example -50 and -30 msec) that are separated by two VOT intervals and are therefore represented half way between these stimuli. This in contrast to the position of the points of the identification functions, which correspond with the actual stimulus on the VOT continuum.

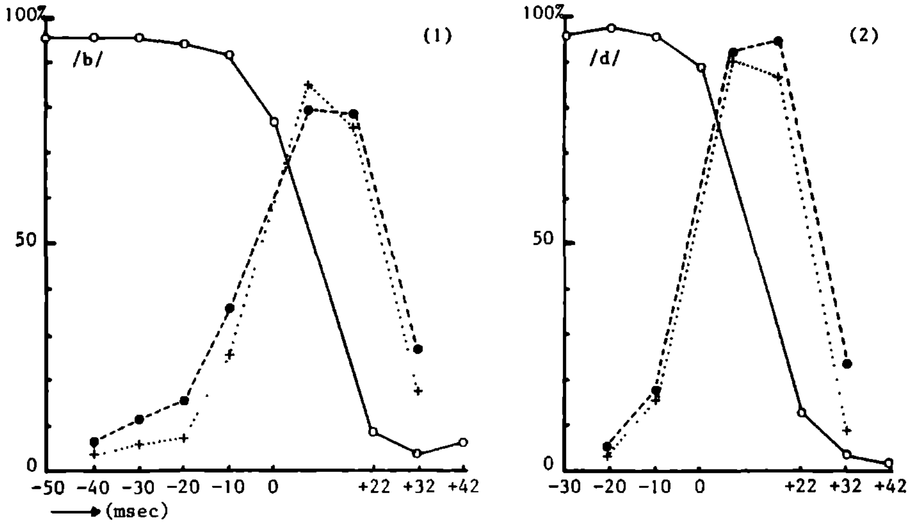

Figure 5: ○——○ identification curves of the [+voice] responses of the /b-p/ (1) and the /d-t/ (2) series, corresponding with the data in table 3.
●----● two-step discrimination curves based upon the observed values.
+·····+ two-step discrimination curves based upon the predicted values.

The identification curves from both series show a perceptual jump from the category [+voice] to the category [-voice], which occurs between 0 and +22 msec VOT. To check whether the acquired responses that correspond with the curves preceding 0 and succeeding +22 msec VOT indeed belong to two discrete categories, the 1% confidence limits for the alternative responses 'voiced' and 'voiceless' were established by means of the binomial-approximation of z that was corrected for continuity. The observed judgements which exceed these limits are considered to belong to discrete categories (Siegel, 1956). With 960 combined responses per stimulus and under the hypothesis that no separate categories would occur during the labeling tasks, it will be evident that stimuli which are judged 54% or more as voiced can be considered as being perceived as [+voice] and stimuli which 46% or less of the time are judged as voiced are considered to belong to the [-voice] perceived category. Stimuli which have a percentage of [+voice] or [-voice] judgements which is between these two limits can be considered as ambivalently perceived.

Referring to table 3, it is evident that indeed the stimuli 1 through 6 of the /b-p/ series can be counted as belonging to the [+voice] category and the stimuli 7 through 9 as belonging to the [-voice] category. Of the /d-t/ series the stimuli 1 through 4 belong to the [+voice] category and the stimuli 5 through 7 to the [-voice] category. No ambivalent perceived stimuli are found in either series.

In figure 5, the discrimination curves of the obtained two-step comparisons of both series show a trough for the intracategory-comparisons and a peak for the intercategory-comparisons.

Under the hypothesis that each two-step comparison would be judged 50% as 'same' and 50% as 'different', which also would imply that no discrete perceptual categories in the stimulus series existed, the 1% limits of confidence for placing the comparisons into two discrete categories were established with the help of the z-approximation of the binomial distribution that was corrected for continuity (Siegel, 1956).

The results of the z-approximation, performed on 768 pooled judgements for each comparison, established that 55% or more 'different'-responses for a comparison indicate that the two relevant stimuli were perceived as being different and that 45% or less 'different'-responses indicate that they were perceived as being the same. A percentage of 'different'-responses between 45% and 55% indicate that a comparison was ambivalently judged and was perceived neither as 'same' nor as 'different'.

Referring to table 3 and figure 5 it appears, then, that the stimuli of the observed intracategory comparisons were perceived as being the same and the stimuli of the observed intercategory comparisons as being different for both the /b-p/ and the /d-t/ series. There were no ambivalently discriminated comparisons.

Finally, for each two-step discrimination the observed and predicted discriminations were compared with each other with the help of the established 1% confidence limits of discreteness of response. On the grounds of the determined limits of 55% and 45%, it appears from table 3 and figure 5 that each of the respective obtained and predicted two-step comparisons from the /b-p/ and the /d-t/ series belongs to the same response-category, being either perceived as 'same' or as 'different'.

The above mentioned findings indicate that the two VOT series were perceived in accordance with the assumptions that were posed with respect to categorical speech perception.

3.4.2. The place of articulation-effect.

Lisker and Abramson (1964) found that the VOT phoneme boundaries of related stop consonants moved systematically in the direction of the burst portion on the VOT continuum when the place of articulation changed from back to front. This phoneme boundary shift in function of the place of articulation could be shown in the production as well as the perception of related stop consonants (Zlatin, 1974; Klatt, 1975; Miller, 1977). Contrary to the systematic difference that exists between phoneme boundaries, the size of the phoneme boundary zones appeared not to be dependent upon the place of articulation and neither in the productive nor perceptual sense to differ systematically from each other (Zlatin, 1974).

To test the place effect, a paired t-test was conducted on the differences which exist between the 50% cross-over scores of the /b-p/ and /d-t/ series (Ostle, 1963; Silverman, 1977). The paired t-test is based on the distribution of the mean difference of an x number of pairs. This mean difference, $\bar{D}$, has a Students t-distribution. By applying the test, it was checked whether the mean difference $\bar{D}$

was significantly different from a difference equal to 0; in other words, the question was whether 0 fell within the 95% confidence limits of a theoretical distribution of $\bar{D}$. First a test was performed on the 50% cross-over differences for each (sex-age) subgroup under the nulhypothesis $\bar{D}= 0$. No difference could be found between the 50% cross-over values for any of these subgroups (p varied from .07 to .92). However, a test of the 24 pooled 50% boundary differences resulted in a significant difference (t= 2,42 df= 23 p= .02). Because the mean differences $\bar{D}$ of the 50% cross-over values corresponding with the data in table 2 are 8,32 msec for the /b-p/ series and 10,42 msec for the /d-t/ series, it can be concluded that there exists indeed a small but significant place of articulation effect. According to the expected tendency the phoneme boundary of the bilabial series lies nearer to the burst portion on the VOT continuum than the boundary of the alveolar series.
We have already referred to the consequences of a varied and place of articulation dependant critical stimulus characteristic of the voicing feature, in connection with a possible transfer of learned information to another perceptual context.
Finally, a paired t-test was conducted on the 24 pairs from the /b-p/ and /d-t/ phoneme zones. The results of this test indicate that the phoneme boundary zones do not differ significantly from each other ($\bar{D}$= -0,50 df= 23 t= -0,69 p= .95).

CHAPTER 4

AUDITORY TRAINING OF THE VOICING DISTINCTION

4.1. PROGRAMMED AUDITORY TRAINING.

As Ling (1978) has recently stated, there is no comprehensive theoretical framework based on fundamental experimental studies of the acquisition of specific auditory capacities.

In our introduction it was pointed out that many published methods of auditory training were based on principles of learning such as programmed instruction and assumed that the training of non-speech sounds facilitated the acquisition of speech sounds. Apart from our conviction that training with non-speech sounds can be benificial for certain non-speech-bound cognitive dimensions (Crul, 1977b), there is some doubt as to whether it would be also beneficial in the case of inadequate phoneme perception because such noises differ considerably from speech sounds as far as their acoustic properties are concerned and may be processed in a different way.

In this research programmed auditory training is applied with the help of speech stimuli in which a critical acoustic cue is held under control, while this cue forms an integral part of the stimuli.

Programmed auditory training has been applied before in the training of children with auditory handicaps. Doehring et al. (1967) have shown that programmed instruction in the case of aphasic children led to an improvement in the distinction of meaningful environmental sounds. This was also shown to be true for children with hearing deficiencies (Saleh, 1965; Doehring, 1968; Crul, 1971).

Improvement in the distinction of speech sounds using automatic programmed instruction with children suffering from speech defects has been reported by Holland and Matthews (1963), Winitz and Preisler (1965), and Winitz (1969). Similar results with aphasic children have been reported by McReynolds (1966), and with children with hearing deficiencies by Amcoff (1968), Ling and Doehring (1969), Doehring and Ling (1971) and Crul (1977a).

Most of the above studies, however, provide limited opportunity for comparison with the present experiment since stimuli were used whose acoustic parameters were not controlled or were unknown, while in many cases discrimination learning was based on the perception of holistic patterns or very crude acoustic differences without taking into account their perceptual relevance. One general conclusion which can be drawn from such studies is that programmed auditory training is very effective. But the conclusion that programmed training might be a very effective method in the acquisition of sound discrimination hardly tallies, as far as speech sounds are concerned, with the finding that generalizing to other contexts cannot be demonstrated, as we also found with, inter alia, the holistic method (Crul, 1977a).

It seems as if such programs often achieve context-bound discriminations which have little relation to phoneme perception in the sense of a restructuring of the varied acoustic signal into the perception of a fixed linguistic unit within any varied context. Ling and Doehring (1969) and Doehring and Ling (1971) were also unable to show any transfer of learning from trained speech stimuli to the same stimuli in untrained contexts in the case of listeners with hearing problems.

In contrast to the negative findings concerning transfer of learned speech discrimination among auditory disordered subjects, positive results have been reported in similar experiments with chinchillas (Kuhl & Miller, 1975, 1978). These mammels do show evidence of transfer of learning of voicing information to a different place of articulation than the one learned as well as to different vowel contexts.

The training of the auditorily handicapped with speech stimuli in which the cues critical for phoneme recognition are controlled is still in an early stage. The reasons for this situation are, inter alia, technical problems involved in manipulating these cues and the fact that most of these cues have only been relatively recently discovered.

It was relatively early on that Wedenberg (1951, 1954), a member of the circle round the phonetician Fant, reported some succes in training children with auditory handicaps with the aid of partially controlled stimuli. Wedenberg's method was to match the speech stimuli to the loss of hearing of his subjects by making use of the basic properties of the speech stimuli, i.e. the frequency-intensity distribution of the formants. He used normal speech stimuli and did not employ programmed instruction.

Bennet and Ling (1977) phonetically examined the critical VOT cue in normal speech stimuli by means of which children with severe hearing difficulties have been successfully trained to distinguish between voiced and voiceless consonants. They utilized programmed discrimination training.

Owens (1978) used systematically obtained feature confusions from normal speech stimuli with auditorily disturbed subjects in an auditory training procedure based on successive approximation, stimulus fading and shaping which was highly specific and geared to each individual.

Lane and Moore (1962) have already been mentioned in the context of their attempt to bring about a categorical perception of the voicing distinction in an adult aphasic patient by means of synthetically produced VOT stimuli using systematic operant discrimination training. However, they did not try to trace the effect of successful reconditioning by means of study of transfer of learning.

Kruger et al. (1972) and Fourcin (1976) have tried to gain an objective insight into the perception of critical cues within the speech signal using manipulated speech stimuli with auditorily handicapped subjects. They suggest that highly specific training would be appropriate deriving from similar findings with the help of computer-controlled stimuli.

4.2. THE EXPOSURE PHENOMENON.

In our research work we have applied operant auditory discrimination training not only because this method appears to be fairly effective in learning to recognize sounds, but principally because of a preliminary finding which we call the exposure phenomenon. This term has been borrowed from Fry et al., who assume that most auditorily handicapped children gain sufficient experience to learn to understand spoken language through use of hearing aids from a very early age and consequently a relatively adequate exposure to normal speech patterns (Fry, 1966, 1975; Whetnal & Fry, 1971; Ling, 1978; Ling & Ling, 1978). The fact that a child who is hard of hearing learns to understand speech and learns to speak is in their opinion more a question of timely exposure to speech stimuli than the degree of hearing loss, while all types of speech cues can be made effective even in the presence of considerable loss of hearing.

Preliminary testing of a phenomenon that appears to bear some relation to the opinion that mere exposure to speech stimuli has an effect on the perception of speech, took place during the preliminary stage of the stimulus construction in this research. An attempt was made to effect the perceptual /b-p/ contrast not only in the context of the words /bɑk-pɑk/, we have already referred to, but also in a /be:r-pe:r/ <bear-pear> context in accordance with the same segmentation procedure referred to in chapter 3. However, in this last case use was made of a voice-lead of 30 msec and a voice-lag of 30 msec with respect to the burst portion. Therefore this /be:r-pe:r/ series consisted of seven stimuli.
During the first listening trials with this series in which subjects were forced to make a choice per stimulus between the alternatives /be:r/ - /pe:r/ it appeared that some listeners with normal hearing were nevertheless unable to perceive a distinct contrast or to identify the stimuli that should have belonged to the /b/ category, except in an ambivalent or quasi-random fashion.
Other listeners categorized the series as expected into two clear categories with the phoneme boundary lying in the region between 0 and +22 msec on the VOT continuum. One possible explanation was that several listeners were no longer able to perceive the [+voice] character of the bilabial stop consonants thanks to an artifical reduction in voice-lead from ±80 to 30 msec or less. In other words it appeared that the stimuli with reduced voice-lead lacked the necessary information of [+voice] for such listeners, while this information appeared to be redundant for others.
One of the adults with ambivalent perception of the voiced representatives of the /b-p/ series was subsequently selected for a longer term experiment which consisted of three phases. Phase 1 consisted of three sessions with a time interval of one month between two consecutive sessions during which the /be:r-pe:r/ identification lists were presented without previous information. Each of the seven stimuli appeared twenty times per session. Phase 2 followed one month after the last session of phase 1 and consisted of four sessions again spaced out at intervals of one month,

during which the /beːr-peːr/ identification lists were again presented until twenty responses were obtained per stimulus in each session. First of all, however, immediately preceding each session a tape which had been constructed by computer, with an n=200 random ordering of the two extreme stimuli of -30 msec and +42 msec VOT and an interstimulus interval of 0.5 sec was presented. Pains were taken to see that both stimuli appeared equally frequently in these 200 presentations. The subject was obliged only to listen without responding and was instructed to attempt to discover a difference between the stimuli. Finally in phase 3, with a latency period of one month from the last session of the second phase, the /beːr-peːr/ series was again presented for identification without prior exposure to the extreme VOT stimuli until 80 responses per stimulus were obtained.
This last session was meant to obtain some insight into the degree of long term retention of possible perceptual learning. The results of these three phases, expressed in the percentage of /b/ responses, are presented in table 4.

		VOT stimuli /beːr-peːr/						
/b/response	sessie	-30	-20	-10	0	+22	+32	+42
Fase 1 before exposure	1	30	40	55	60	20	10	0
	2	30	50	30	65	20	20	0
	3	65	40	65	75	25	10	0
	Mean	41,67	43,33	50,00	66,67	21,67	13,33	0
Fase 2 after exposure	4	85	100	100	65	15	15	0
	5	75	90	90	55	30	15	15
	6	85	90	100	85	50	15	15
	7	75	90	100	85	85	10	10
	Mean	80,00	92,50	97,50	72,50	45,00	13,75	10,00
Fase 3 retest	8	90	100	95	75	35	5	10

Table 4: Percentages of VOT stimuli identified as voiced from the /beːr-peːr/ series over eight sessions with the same listener with normal hearing. Every session score in phase 1 and 2 is based on twenty responses and the scores in phase 3 are based on eighty responses. There is an interval of one month between any two successive sessions. In session 1 - 3 the stimuli were presented without previous information and in sessions 4 - 7 after a random exposure (n=200) of the extreme -30 and +42 msec VOT stimuli. In session 8,which took place one month after the last session in phase 2, the stimuli were once again presented without previous information. The results of the third phase can be interpreted as a measure of the retention.

A first glance at table 4 shows that for each session of phase 1 the first three stimuli, which correspond to a voicing-lead in the speech signal, were not identified as clearly belonging to a certain voicing category, while this was indeed the case for every session of phase two as well as in the retest. In order to examine these differing response trends in the identification functions of the three phases more closely the results of table 4 have been subjected to three analyses.

First of all the three identification functions in phase 1 were studied to see if they really differed from each other. In view of the fact that the curves are the result of the responses by the same listener to the same seven stimuli, a non-parametric test for related samples was carried out on the absolute response scores. On the basis of a Friedman two-way analysis of variance (Siegel, 1956) the question as to whether the response scores of the seven stimuli over the three sessions differed from each other appears to be answerable in the negative ($\chi_r^2 = 2.57$; df= 2; two-tailed $p > .20$). One may thus conclude that there is no significant difference between the three identification functions of the first phase.

The same test procedure leads us to conclude that the four identification curves of the second phase also do not differ significantly from each other ($\chi_r^2 = 2.44$; df= 3; two-tailed $p > .30$).

These findings lead us to combine the scores of the three sessions in phase 1 which results in one single pre-exposure identification function. The scores of the four sessions in phase 2 were dealt with in the same way, so that one single post-exposure function was obtained. These functions are shown in figure 6 together with the retest results.

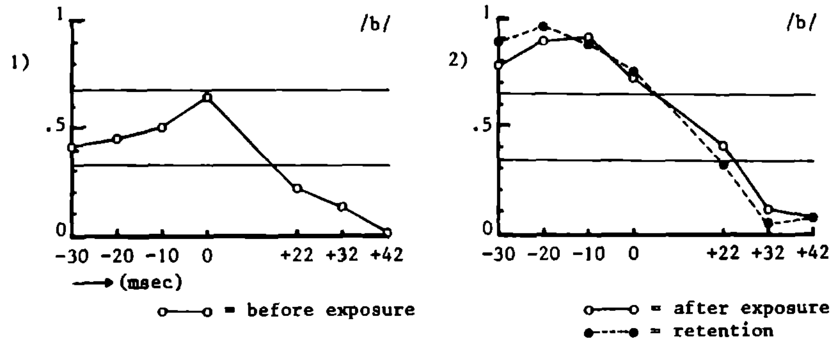

Figure 6:Identification functions based on the average percentage of [+voice] responses of the stimuli from the VOT series /beːr-peːr/ per phase. See also table 4.
Figure 6(1) is the function of the responses prior to random exposure to the two extreme VOT stimuli. In figure 6(2) the functions of the responses after exposure to the extreme stimuli and of the retest scores are shown. The graphs show the exact two-tailed 1% confidence levels of the responses by two horizontal lines.

Just as table 4 suggested, the functions given in figure 6 show a clear difference between the pre- and post-exposure functions.

Before determining the effect of the exposure it is relevant to mention the subjective experiences of the listener with respect to the different stimuli. During 3 /b-p/ identification sessions in the first phase, the only clearly identified initial consonant was a /p/. But during the first random exposure to the two extreme VOT stimuli in the second phase a distinction was quickly perceived between a harsh /p/, with a VOT value of +42 msec, and a /p/-variant, with a VOT of -30 msec. The latter was later described as 'mellow', but was never identified as a normal /b/ phoneme. The distinction between the 'subjective' features /harsh-mellow/ was also manifested in the identifications, which followed the exposure to the extreme stimuli, as can be concluded from the altered response gradient in figure 6(2). The listener never heard the /b/ phoneme in the word /be:r/ during the identification tasks, but, apparently as a result of the exposure to the two extreme stimuli, he could distinguish between two /p/-variants. Admittedly these allophones never occurred as relevant linguistic contrasts in the speech environment of the listener, but they could be categorically described with the subjective features /harsh/ versus /mellow/. Actually it appeared that the listener who, before the experiment, only perceived the normal voicing contrast between /b/ and /p/, was able during the course of the experiment, to identify a third VOT category. He had learned on the basis of VOT differences to perceive a three category system, consisting of /+voice/, /mellow p/ and /harsh p/ in place of the two-category feature system derived from his linguistic environment, consisting of the features /+voice/ and /-voice/.

As was shown in the study of Lisker and Abramson (1964), elsewhere there are, indeed, three-category languages, such as is the case with the voicing contrast in Thai.

The testing of the identification scores of the first and second phases with the binomial test indicates that the effect of the exposure procedure, indeed, concerns a change from an inadequate categorizing of a stimulus continuum to a more categorical differentiation between respective stimuli. The exact levels of significance (at the 1% level, two-tailed) of the chance area of the seven identification scores were determined with the z-approximation of the binomial test that was corrected for continuity (Siegel, 1956). Considering the identification curve of the first phase in figure 6(1) these limits lie at 67% and 33% for a combined score of N= 60 for each stimulus. Considering the perceptual functions in the second and third phases, the limits lie at 65% and 35% for a score of N= 80 for each stimulus. In figures 6(1) and 6(2) the upper and lower limits are represented as horizontal lines which cut the axis at the correct point. Here it also appears that a good /p/ category was perceived in the first phase but that the relevant stimuli of the expected /b/ category, corresponding with the VOT values from -30 msec to 0 msec, were answered quasi randomly.

The identification functions clearly show in figure 6(2) the appearance of the /mellow p/ category after exposure to the extreme VOT values, because the stimuli which correspond with the VOT values from -30 msec to 0 msec have left the chance area and have formed a separate category which contrasts significantly with the /harsh p/ category.

Although the new function exhibits a relatively flat slope and the obtained categorizing does not implicate an ideal categorical perception, it can be concluded that the enhancement of the /p/ contrast is brought about solely on the basis of an exposure to the two extreme VOT stimuli.

The retention function shown in figure 6(2) which is based upon the labeling scores achieved a month after the last session of the second phase, without the presentation of the extreme stimuli before the labeling task, appears almost identical to the perceptual function of the second phase. It can be concluded that, at least in the experimental situation, the normal listener permanently judged the features /harsh/ and /mellow/ to be two perceptually relevant units, once he was aware of such a distinction.

Finally, the changes in the identification functions of the three phases could also be expressed in the Categorical Score (C), a modification of a categorical measure developed by Fourcin et al. (1978) and Simon and Fourcin (1978). This measure indicates the degree to which stimuli from a continuum are responded to in a categorical mode. With this measure, the responses are expressed in a number that can vary from -1.0 to +1.0. The C-score is dependent upon the degree to which all the items on one side of the phoneme boundary are labeled as category x and on the other side as category y. In the normal categorized /be:r-pe:r/ series, the phoneme boundary fell as expected between the stimuli with a Vot value of 0 msec and +22 msec respectively. Thus, the four stimuli with a VOT which varied from -30 msec to 0 msec could be considered to belong to the same category, while the three stimuli with a VOT from +22 msec to +42 msec could be considered to belong to the contrasting category. In view of the fact that random response means that 50% is correct, the Categorical Score (C) can be defined as follows:

$$\text{Discrimination} = D = \frac{\text{(number correct identifications)}}{\text{(total possible number correct identifications)}}$$

$$\text{Categorical Score} = C = 2(D - 0.5)$$

C is 0 for chance response. A negative C refers to a reversal of the perceptual function. A sloping function has a C of 0.5, while a C above 0.75 indicates a significant abrupt, categorical response function.

The C-score and the related perceptual categories (i.e. chance response, reversal, sloping function and categorical response) will be used regularly in the course of this research.

The C-score for the identifications of the first phase was 0.33, for the second phase 0.64, and for the perceptual retention function 0.74. Thus there occured a positive change from an inadequately perceived stimulus contrast to an almost significant categorically perceived contrast.

We can summarize a number of facts about the exposure phenomenon which are essential for the continuation of the research.

1: It is possible through only active listening to two extreme, contrasting stimuli from a continuum to bring about a perceptual contrast that did not exist before. An explanation for such a perceptual shift can be found in the opinion of Eiler et al. (1979) about the influence of linguistic experience on learning to discriminate speech sounds and of Ades (1977) about the characteristics of the stimuli and the task in the exposure experiment.

2: For the perceptual acquisition of a phoneme contrast, it does not appear to be necessary to refer to ones own speech. In other words, a perceptual contrast can be brought about through listening alone, without a vocal or subvocal speech response. In this case, the contrast was not even part of the articulatory repertoir of the listener, so reference to it was not possible. We do not claim that the speech mechanism can not play an important or facilitating role with respect to the learning of auditory speech discrimination but we do show that it is not necessary, per se, to involve it in the perceptual learning process. This can be of consequence in the case of listeners who are still incapable of realizing certain phoneme contrasts in their speech. It argues, further, for the classical viewpoint of speechtherapists, namely, that before beginning a training program for inadequate articulation, it is desirable to offer discrimination training for the relevant speech sounds (Van Riper & Irwin, 1958).

3: The learned distinction need not always agree with the feature system of the linguistic environment of a speaker-listener with normal speech and hearing abilities. A similar perceptual contrast can be subjective in nature and can, consequently, have a subjective linguistic relevance for the speech perception of a normal or hearing impaired listener.

4: With respect to the learning or improving of perceptual contrast, the perceptual function from the first and second phases (see figure 6) can be interpreted as a response gradient. Here one sees a stimulus generalization of the trained -30 msec stimulus over the physically nearby stimuli, respectively those with a VOT value of -20, -10 and 0 msec. This primary stimulus generalization of a certain stimulus over an entire, separate category was mentioned in the early study of Lane and Moore (1962). It is of great importance with respect to the predicted effect of training with selected stimuli from our /b-p/ and /d-t/ series.

5: These and the previously mentioned facts from earlier studies which indicate that it is effective to teach auditory discrimination by means of an operant training procedure led to the choice in this study of a simple programmed instruction in the form of operant conditioning as teaching method. It was expected that the actual listening to selected speech stimuli together with an active (nonspeech) response behaviour and systematic response consequences in the form of immediate knowledge of results as well as stimulus repetitions would make the learning to distinguish perceptual contrasts yet more effective.

4.3. THE TRAINING OF THE VOICING CONTRAST.

4.3.1. Selection of the hard-of-hearing experimental and control subjects.

Deaf speakers often fail to realize certain time relations in their speech patterns that normal speakers do automatically (Kruger et al., 1972). Thus voiced and voiceless cognates of consonants are frequently confused because insufficient auditory feedback takes place during the act of speaking (Calvert, 1962; Gemmil & John, 1976). The hard-of-hearing also fail to differentiate, as clearly as a normal speaker, between voiced and voiceless stop consonants (Gilbert & Campbell, 1977). A normal speaker generally pronounces voiced and voiceless initial consonants within relatively clearly separate categories on the VOT continuum and enlarges this separation when speaking emphatically (e.g. citation-speech) or when producing isolated pronounced minimal pairs (Lisker & Abramson, 1967). It has been shown, for example, that mothers when speaking with their young children consciously maximize the phonetic VOT contrast (Baran et al., 1976).

In the case of the hard-of-hearing, one often observes only a slight separation and frequently even an overlap of the voiced and voiceless produced categories on the VOT continuum. In contrast with those with normal hearing, they make a less clean distinction between two related articulations. With respect to the VOT separation of voiced and voiceless relatives, one observes a 'perceptual redundancy need' in the hard-of-hearing.

A similar phenomenon is the 'blurring distinctiveness' in sloppily pronounced, fluent speech, where an instability of contrast relations and a reduction of physical (VOT) distances between categories can be seen (Lisker & Abramson, 1967).

We mentioned in section 2.3. that Bennet and Ling (1977) were able to effect a greater VOT separation as well between the perceived as between the produced categories in children whose hearing was severely impaired by means of perception training.

Despite the problems concerning VOT perception and VOT production it appears that a major source of information for the nearly deaf is the voicing distinction in speech perception.

Concerning the importance of the perception of the feature of voicing in those with hearing disabilities, one can give several explanations. Boothroyd (1978) remarked that the most noticeable effect of sensori-neural deafness is the loss of the auditory capacity to perceive acoustic cues but that this problem can be partially alleviated by sound amplification. As is also true of Fry (1975), he assumes that the people with a sensori-neural hearing loss use the same acoustic cues in speech perception as those used by people who hear normally. Certain cues and their associated distinctive features seem, however, to be more resistant to transmission problems than others. One of the most stable auditory cues appears to be that on which is based the perception of the voicing distinction, while, for example, the auditory information needed to perceive the place of an articulation feature is readily obscured by interference.

The strong resistance of the voicing feature to varying forms of distortion -for instance the systematic filtering out of frequencies relevant to speech which can be compared with certain hearing disabilities- is demonstrated in the feature confusions of the listening experiments carried out on subjects with normal hearing by Miller and Nicely (1955) and later by Singh (1971) and Binnie et al. (1974) (see also Campbell, 1974).

Besides these experiments in which the redundancy usually present in normal speech signals was artificially reduced, Danhauer and Singh (1975) were able to demonstrate the great resistance of the voicing feature by means of multivariate aposteriori analyses of the perception of distinctive features in subjects with various hearing disabilities. According to their analyses, it appears that this particular feature is a primary source of auditory information in the speech perception of very hard-of-hearing with only residual hearing in the lower speech frequencies. But the importance of this feature seems to be somewhat diminished in, for example, those hard-of-hearing still capable of perceiving the higher speech frequencies. Those with severely impaired hearing generally manifest their own perceptual strategy. On the one hand, they make use of the auditory information contained in the visually imperceivable voicing contrast, while, on the other hand, they supplement the information by means of lipreading such cues as the place of articulation which generally can not be auditorily perceived (see also Campbell, 1974).

The stability of the voicing feature in the perception of subjects with impaired and normal hearing is also demonstrated with aposteriori analyses by Bilger and Wang (1976), Wang and Bilger (1978) and Stewart et al. (1979).

It is known that, in the case of severe sensori-neural hearing disturbance, and generally with deafness, as well, one finds residual hearing in the lower speech frequencies which plays an important rôle in the perception of the voicing cue. Erber (1972a) demonstrated that the deaf, whose hearing loss extended to the lower speech frequencies, as well, were only able to perceive rough time-intensity cues, based upon the envelope contours of his stimulus words. He contended that this perception was most likely of a vibro-tactile nature. He found, moreover, that a severely handicapped with a hearing threshold that did not exceed 90 dB could perceive the voiced-voiceless distinction based upon residual hearing in the lower frequencies (1972b, see also Boothroyd, 1978). Such an extensively disabled subject receives generally inadequate acoustic information concerning the manner and the place of articulation. Entirely deaf people fail to hear even the voice contrast and generally give quasi-random voice responses on identification tests with voiced-voiceless items (Fourcin, 1978).

Ling (1978) contended that the inability of the severely hard-of-hearing to distinguish between voiced and voiceless cues was usually caused not by insufficient hearing potential but by a poorly fitted hearing aid or by excessive attention paid to visual attempts towards speech perception to the detriment of the use of auditory cues.

The above data point up two important aspects of the perception by the hard-of-hearing of the distinction between voiced and voiceless features. Firstly, the loss of hearing must be very great, more than 90 dB over the entire speech frequency range, in fact, to result in severely impaired perception of this contrast. Secondly, the slope of the hearing threshold is of importance, in view of the fact that residual hearing in the lower speech frequencies is sufficient for the perception of this contrast.

The hard-of-hearing subjects in this learning research were selected from children in a school for the hard-of-hearing and not from a group of deaf children. Because of this, there was some difficulty in finding subjects who, because of their hearing, were not capable of separating the stimuli of the /d-t/ and /b-p/ series into two categories.

Originally, a total of 30 prelingual hard-of-hearing children of about 10 years of age with varied, two-sided sensori-neural hearing loss received the /d-t/ label series. The procedure and the material were the same as described in section 3.3. With the help of the probe-items before the stimulus lists, the most comfortable intensity level for the presentation of the lable stimuli was individually determined. Use was made of normal linear audio-amplification. Preference was given to linear amplification over individual hearing aids because the former, unlike the latter, delivers the total possible range of speech sounds (Ling, 1978).

The /d-t/ identification series was presented twice after which the identification function with n=20 responses for each stimulus was established. It turned out that 19 of the 30 children showed significant deviations in many aspects of the placement and the form of the graphs of their functions relative to the established norms of table 2. Three children gave quasi-random responses to the stimuli or perceived all stimuli as primarily voiced. These three made absolutely no categorization. Sixteen children did categorize but showed abnormal boundary values in the phoneme boundary area in favor of one of the two categories. In several cases, it appeared that the separation between the categories transgressed the permitted norm. In agreement with the results in the literature, it was observed that the most deviant, quasi-random or on one phoneme fixed responses came from children with very severe hearing-loss, where relatively little hearing capacity remained in the lower speech frequencies. Extremely steep hearing thresholds and hearing losses of less than 90 dB often resulted in a boundary shift between the two categories.

Of the 19 hard-of-hearing children with abnormalities in the identification function, eight were selected who were considered to be capable to take part in a long listening research program. Of the eight, four were taken as a control group and four as an experimental group. The control group was only used in the transfer research. More detailed data concerning the subjects are given in table 5. As shown, subjects 1 and 2 were relatively light hearing impaired, subject 3 was severely hard-of-hearing and subject 4 was on the boundary between extremely severe disability and deafness.

		subj	sex/age	sensorineural hearing loss Hz 250 500 1000 2000 4000	etiology	hearing aid	speech	HAWIK IQ	schoolresults
EXPERIMENTAL GROUP	hearing impaired	1	boy / 11	R 10 30 30 50 50 L 20 30 40 50 50	hereditary	+	lightly disturbed	V 109 P 122	normal
		2	boy / 13	R 20 30 40 40 30 L - - - - -deaf- - - - -	unknown	+	lightly disturbed	V 92 P 111	normal
		3	girl/ 12	R 70 80 75 70 70 L - - - - -deaf- - - - -	prelingual meningitis	+	intelligible, heavy disturbed	V --- P 127	normal
		4	boy / 12	R 60 80 90 95 110 L 60 80 90 90 110	prematurity asphyxia	+	typical deaf speech	V 70 P 124	normal
	speech disordered	5	girl/ 11	R - - - - normal - - - - L - - - - normal - - - -	unknown	-	development dysphasia	V 60 P 99	learning disorders
		6	girl/ 11	R - - - - normal - - - - L - - - - normal - - - -	unknown	-	development dysphasia	V 74 P 86	learning disorders
		7	girl/ 10	R - - - - normal - - - - L - - - - normal - - - -	prematurity	-	development dysphasia	V 75 P 97	learning disorders
CONTROL GROUP	hearing impaired	8	boy / 10	R 20 30 30 40 45 L 20 30 30 40 45	unknown	+	good	V 94 P 111	normal
		9	girl/ 12	R 10 10 30 60 60 L 10 20 40 60 60	prematurity asphyxia	+	lightly disturbed	V 94 P 107	normal
		10	boy / 11	R 50 45 50 75 80 L 35 30 40 30 40	prelingual meningitis	+	lightly disturbed	V 90 P 108	normal
		11	boy / 12	R 0 20 50 70 50 L 0 30 50 70 90	prematurity difficult delivery	+	lightly disturbed	V 90 P 99	normol

Table 5: An overview of the experimental and control subjects. The hearing thresholds were for the relevant speech frequencies, per octave from 250 to 4000 Hz, obtained with a Peters AP31 audiometer for both ears and are expressed in dB (norm ISO R389 1970). The HAWIK-intelligence test is a European variation of the WISC. It contains both a verbal (V) and a performance (P) section. General school achievement levels are relative and are only mentioned with respect to the possibilities of a child who attends a special school for the hard-of-hearing or for the speech-disabled.

After the selection, the eight subjects were given in daily sessions the /d-t/ and /b-p/ label and two-step discrimination lists until, for every listener, 40 identifications per stimulus and 32 comparisons per discrimination pair was achieved. The respective perception functions based on the first 20 identifications per stimulus and on the next 20 responses per stimulus were inspected for response consistency by checking that the curves had more or less the same graph. Such consistency was, in fact, observed in the case of each subject.

The final boundary values of the identification functions, which were based upon the [+voice] responses, were interpolated linearly and are to be found along with the corresponding norms in table 8 of chapter 5 (page 72). In the same table one can find the results of three speech-impaired experimental subjects (see section 4.3.2). To facilitate quick comparison, the post training results can be found there,as well. In figure 7, also in chapter 5 (page 71), one sees the identification functions of the [+voice] responses together with the discrimination curves of the four hard-of-hearing and three speech-impaired experimental subjects, before and after the training.

4.3.1.1. Analysis of the categorization and determination of the selective training stimuli for the hard-of-hearing subjects.

From the identification functions from before the training shown in figure 7 and the boundary values from table 8, it appeared that for the hearing-impaired subjects different training tactics were called for. It was the case, namely, that subjects 1 and 2 required a realignment procedure while subjects 3 and 4 needed an enhancement procedure. Subjects 1 and 2 categorized the /b-p/ series well, as can be seen from the C-scores listed in table 10 (see chapter 5, page 74). The first subject achieved a C-score of 1.00 while the second received a C-score of .96. The responses on the /d-t/ series approximate a significant categorizing which can be seen from the C-scores of .71 and .69. The entire phoneme boundary region, however, is shifted in the direction of the voiceless category. From the identification functions given in figure 7, one sees, that the two neighboring stimuli, corresponding with the VOT values of 0 msec and +22 msec were both labeled as belonging to the [+voice] category instead of [+voice] and [-voice] , respectively. In these two cases it would be of little value to train with the extreme VOT stimuli from the /d-t/ series, considering the fact that these extreme stimuli are already perfectly perceived as contrasting. The realignment procedure for these two subjects consisted of an attempt to effect a boundary shift in the direction of the voiced category by means of training the subjects to identify the stimuli (VOT = 0 msec) and (VOT = +22 msec) as two contrasting cognates.

From the high discrimination scores, before the training, between the stimuli (VOT = -10 msec) and (VOT = +22 msec) one can deduce that subject 1 was capable of perceiving the acoustic differences between both stimuli but that he placed them both in the same category when it came to absolute identification. Thus, once again, it appears that acoustic discrimination does not imply that the stimuli are placed in separate phoneme categories. Likewise, one can not always predict identification based upon discrimination (cf. Simon & Fourcin, 1978).

As training stimuli in the learning research for the first two subjects, the neighboring stimuli (VOT = 0 msec) and (VOT = +22 msec) were used.

Subject 3 categorized the /b-p/ series well and achieved, consequently, a C-score of .96. The /d-t/ series was, before the training, not categorical perceived and the C-score of .27 strongly suggested a random answering behavior, even though a slight trend toward categorizing can be seen from figure 7 (page 71). The function assumes the required values of 75% and 25% at no point and a boundary zone is not distinquishable (see table 8, pag. 72). Furthermore, the absence of a clear peak in the two-step discrimination curve suggests the inability to perceive the required differences between the stimuli at the correct position on the VOT continuum. The identification function of subject 3 required an enhancement procedure for both the /d/ as the /t/ categories.

In the above (see chapter 4.2) it has been assumed, amongst other reasons because of the findings of Lane and Moore (1962), that an effective discrimination training with extreme stimuli, derived from a non categorically perceived VOT continuum, would have a generalizing effect on the stimuli of the continuum near to the training stimuli. As a result, two distinct perceptual categories could be created. The response gradient of figure 6 which was established after the exposure procedure serves to support this assumption.

Another argument for the use of the extreme stimuli in the case of a labile categorizing of one or both categories in a series of stimuli can be based upon the recent discoveries of Simon and Fourcin (1978). Their results indicate that young English children of 2 and 3 years of age perceive only the extreme VOT stimuli from a continuum in an adequate fashion. This is apparently the case because these stimuli were the most closely related to those naturally produced in an English speaking environment. These children would not yet have a general-conceptual representation of a consonant phoneme and, consequently, would categorize such a phoneme in all its extreme contrasting redundancy quite well. They would not be able to do this, however, for the artificially constructed allophone variations of this phoneme.

The experimental subjects in our research who displayed a labile categorizing of the VOT continuum showed equally few signs of having a general conceptual representation of the consonant allophones of /b/, /p/, /d/ or /t/. It seemed, thus, the most effective approach to try to develop the desired contrast by means of the most VOT contrasting or separated stimuli irrespective of whether these agree with those with are most naturally produced in the Dutch language (for the use of maximal feature differences in discrimination learning see also Gibson, 1969).

The third experimental subject received a training which made use of the extreme stimuli (VOT = -30 msec versus VOT = +42 msec) of the /d-t/ series.

The fourth and most severely handicapped subject was capable of categorizing the /b-p/ series to some extent before the training. There is, however, in a case such as this, a substantial redundancy need, apparently as a result of the hearing loss, which produces an exceptionally large separation between the categories and yield an identification function whose graph is relatively flat. This is expressed by a C-score of .50 and clearly shown in the function given in figure 7. The stimuli of the /d-t/ series were answered in a quasi-random fashion by the fourth subject. Apparently no difference was heard between the seven stimuli during the identification test by this subject. This quasi-random answering was expressed by a C-score of .10. This subject also made an adequate discrimination between the relevant stimuluspair (VOT = 0 msec and VOT = +32 msec). This adequate discrimination does not imply, however, that the two stimuli are identified as different.

The training for the fourth subject consisted of an enhancement procedure that tought the /d-t/ contrast with the extreme stimuli (VOT = -30 msec and VOT = +42 msec). Further it was checked wheter the possibly effective training involving the /d-t/ contrast could have a positive effect on the deviant perception of the /b-p/ contrast in the sense that the separation between the categories would become smaller.

This would manifest itself in a sharpening of the /b-p/ identification functions.

4.3.2. Selection of the speech-impaired experimental subjects.

For the purposes of this research, three children with developmental dysphasia were chosen from a group of 10 normally hearing, but speech impaired children of about 10 years of age. These children were enrolled in the department for those with speech impediments in the same school from which the hard-of-hearing subjects were selected. They all suffered speech discrimination problems according to the Auditory Discrimination Test (ADIT, Crul & Peters, 1976), a Dutch version of the American Wepman Test of auditory discrimination (1958) and the test of Pronovost and Dumbleton (1953). Moreover, nine of these ten children could not adequately categorize the /b-p/ or /d-t/ VOT series. The procedure for obtaining 40 lable and 32 discrimination responses per stimulus of the /b-p/ and /d-t/ series from the 3 speech impaired subjects was the same as that used in the case of the hard-of-hearing subjects.
According to the criteria of Benton (1964, 1978) and Ingram (1975), the speech impediments of the three children exhibited the so-called developmental speech disorder syndrome which is generally called developmental dysphasia. These criteria require that the speech disorder not be a result of a mental defect, a severely impaired hearing (according to standard-audiometric criteria), a severe neuro-motor disability nor a deep disturbance in the interpersonal relation. The main symptoms consist of a reduced capacity to perceive and understand oral language and, above all, a diminished ability to express oneself verbally. Substantial learning disabilities are often related to developmental dysphasia. A more detailed overview of the three dysphasic subjects is given in tabel 5 (see section 4.3.1).
It is unknown whether or not a causal relationships existed between the problems these children had with speech perception and their speech production. A causal relationship between speech discrimination and articulation has been generally acknowledged by speech therapists (Travis & Rasmus, 1931; Templin, 1957; Van Riper & Irwin, 1958; Cohen & Diehl, 1963; Spriesterbach & Curtis, 1964; Weiner, 1967; Haggard et al., 1971; Irwin, 1974; Monnin & Huntington, 1974). None of these studies, unlike the investigation described here, had categorical perception as a subject. Although it was not the goal of this research to demonstrate a causal relationship between hearing and speech, it is relevant to give some possible explanations concerning the diminished capacity of the subjects with speech impediments to categorize the VOT stimuli correctly.
A theory arising from the domain of otoneurology relates the physical redundancy of the speech signal and the biological redundancy of the central-perception mechanism that serves to process speech. In the case of damage to the auditory-central nervous system at a high level, normal speech is generally perceived adequately because the central nervous system, despite substantial damage and thanks to redundancy in the speech signal, is still capable of processing the relevant information. However, in the case of reduced redundancy in the speech signal those who suffer damage on the central auditory level exhibit a significant disturbance in speech perception. This theory is based upon the results of the so-called low redundancy speech tests where

the speech signal was made systematically more sensitive to disturbance by means of noise or temporal and spectral modifications and distortions (Bocca et al., 1954; Bocca & Calearo, 1963; Calearo & Antonelli, 1963; Lidén, 1964; Maspétiol & Semette, 1964; Palva, 1965; Korsan-Bengtsen, 1973).

Of the stimuli in the /b-p/ and /d-t/ series one notes that the redundancy is diminished. This results either from the reduction of the prevoicing portion, where the separation between the two categories is systematically reduced, or from the elimination of a certain segment from the signal directly after the explosion burst. According to some experimenters, not only the dominating VOT cue, but also the formanttransitions after the burst-release contain important information for the perception of the voicing characteristic of initial prevocal stops (Stevens & Klat, 1974; Lisker, 1978).

As we could demonstrate in our pre-investigation of the exposure phenomenon (see section 4.2), a similar stimulus manipulation also caused problems in the categorization of a /b-p/ series in a listener without a central-auditory disturbance. It is well-known that a symptome strongly related to developmental dysphasia is the diminished capacity to perceive and to understand spoken language adequately. Consequently, it is possible that the abnormal perception in this respect by the three dysphasic subjects in the pre-training phase is due to the reduced redundancy of the manipulated VOT stimuli.

A second possibility which might explain the labile capacity of the speech impaired subjects to categorize the VOT stimuli finds its origins in speech therapeutic research. It has been suggested, namely, that speech perception and speech production influence one another to a great degree. That perception supports production is hardly surprising. That a converse relationship exists, as well, was experimentally supported by Winitz and Bellerose (1963; see also Winitz, 1969). In an earlier study concerning the process of phoneme development in young children, Shvachkin (1973) describes the influence of articulation upon the phoneme perception during speech development. It was observed that sounds not yet mastered by a young child are distinguished later than those sounds that the child can already produce. It consequently seems that a reciprocal interaction exists between perception and articulation. Similar findings do not exclude the possibility that the abnormal speech perception exhibited by the 3 dysphasic subjects was influenced by their impaired articulation.

A third possible explanation of the fact that the VOT stimuli were not adequately categorized by the dysphasic children is given by a relatively recent neuropsychological finding. It seems that patients suffering from developmental dysphasia or an acquired aphasia display a diminished or labile capacity to perceive speech sounds in 'highly coded' form (Blumstein et al., 1977b). Some speech sounds appear to be more coded than others. In acoustical terms this means that they must undergo a great degree of restructuring as a function of nearby phonemes. In psychological terms this could mean that special processing is required in order to perceive them. The phonemes which display the greatest context-conditioned variations, such as the stop consonants, are amongst the most highly encoded speech sounds. Vowels are the least encoded speech sounds and form the lower bound of the encoding range. In the studies in which the coding hierarchy of diverse speech sounds is described, use is made of the dichotic listening experiment as method and of computer synthesized speech stimuli (Day & Vigorito, 1973; Cutting, 1973). Earlier, it has been shown in this way that the perceptual processing of these most highly encoded phonemes occurs specifically in the major hemisphere (Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970; Liberman, 1973b).

From the experiments of Tallal and Piercy (1973,1974,1975,1978) and Tallal and Newcombe (1976), in which synthetic speech signals were also used, one can conclude that the diminished capacity of dysphasic children and adults to perceive coded speech probably results from the fact that these people are unable to process the rapid and short critical acoustic information in the speech signal in a normal way (see also Lowe & Campbell, 1965).

This information concerns, among others, the rapid and short phoneme transitions upon which the perception of the place of articulation of stop consonants rests. The reduced capacity of rapid processing of critical temporal information concerning the VOT perception and voice contrast has not been investigated by these experimenters. But several recent neuro-psychological studies report that the VOT perception in the cases of aphasic and dysphasic patients is also disturbed. Carpenter and Rutherford (1973) found that particularly perception of the temporal cues, the VOT in voiced and voiceless consonants, for example, is not normal in adult aphasic patients. Blumstein et al. (1977a) and Basso et al. (1977) found great deviations in the categorizing of synthetically produced VOT stimuli in the case of aphasic adults with cerebral lesions in the major hemisphere. In the research of Blumstein et al.,it was observed that the performances of some listeners with respect to the VOT stimuli did not exhibit the expected agreement between perceptual identification and discrimination. Sometimes the discrimination appeared to remain undisturbed while correct identification of VOT stimuli proved impossible. On these grounds, it has been suggested that the perception of an aphasic listener with such disabilities remained undisturbed in so far as this took place in the non-speech mode, while the patient exhibited disturbances in the speech mode. The disintegration of the close connection between perceptual discrimination and identification suggests, moreover, that in the aphasic listeners an acoustic property detector is functioning normally along with a linguistic disability to adequately convert an acoustic signal in a stabily perceived phoneme.

4.3.2.1. Analysis of the stimulus categorization and determination of the training stimuli for the speech impaired experimental subjects.

From table 8 and 10 and from figure 7 one can establish that the 5th subject could categorize neither the /b-p/ nor the /d-t/ series. This subject received a C-score of categorization of .28 for the bilabial series and .14 for the alveolar series. The identifications of stimuli in both series tended strongly toward quasi-random responses although there was a slight tendency to identify the stimuli as being voiced. The two-step discriminations between the /b-p/ pairs indicated that absolutely no difference was perceived between the stimuli. The discriminations of the /d-t/ comparisons were not much better. Subject 5 received a training with extreme stimuli from the /d-t/ series with VOT values of -30 msec and +42 msec, respectively. Further, it remained to be seen to what extent a successful training of the /d-t/ series could influence the categorization of the /b-p/ series.

Subject number 6 labeled the voiced stimuli of the /b-p/ series correctly, while the remaining stimuli, corresponding to VOT values of +22 msec through +42 msec, were not identified as voiceless but rather in a quasi-random fashion. This subject received a C-score of .67. The /b-p/ discrimination curve in figure 7 indicate that the expected acoustic differences between the relevant stimuli of the continuum were, indeed, perceived. The /d-t/ categorization appeared to be quite normal for both the labeling (C= .89) and the discrimination. Subject 6 was given a training with the extreme stimuli from the /b-p/ series, i.e. VOT values of -50 msec and +42 msec, with the intention of enlarging the perceptual contrast between the stimuli on both sides of the phoneme boundary.

Subject number 7 gave totally random responses to the label stimuli of the /b-p/ continuum (C= -.03). The discrimination curve tended to indicate a perception of differences between the stimuli on both sides of the objective phoneme boundary. The /d-t/ series was nearly correctly labeled (C= .74), but the discrimination curve showed no clear peak. Subject 7 received also a training with the extreme VOT stimuli (-50 msec and +42 msec) from the /b-p/ series after which the effects on the categorization of the /d-t/ series was examined.

4.3.3. The training.

4.3.3.1. Training stimuli.

The selected VOT training stimuli for the seven subjects are given in table 6.

subject	/b-p/ stimuli	/d-t/ stimuli
1		0 +22
2		0 +22
3		-30 +42
4	transfer	-30 +42
5	transfer	-30 +42
6	-50 +42	
7	-50 +42	transfer

Table 6: selected training stimuli in msec VOT for the seven subjects and the cases in which a transfer from the learned voicing contrast to a non-learned voicing contrast with a different place of articulation is studied.

It appears from this table that the relevant stimuli with which it was attempted to establish an improvement in the [+voice] and [-voice] perception were selected on the grounds of the individual auditory labeling performances of each subject. With the help of a PDP 11/45 computer, each of the stimulus pairs was recorded on tape in the form of four blocks, each block consisting of 20 stimuli in random order. Each of the two stimuli in a pair occured with the same frequency within a given block. The four random blocks were, in turn, presented in a random order during the consecutive learning sessions.

4.3.3.2. Apparatus.

The apparatus used in presenting the training and feed-back stimuli consisted of two Revox A-77 tape recorders with normal lineair amplifiers and two sets of individually adjustable AKG-K150 headphones, one for the experimenter and one which was equipped with circumaural cushions for the experimental subject. These headphones reproduced the signals from both recorders binaurally. One of the tape recorders presented the stimuli from the training tape. This apparatus could be controlled via a pedal which allowed the recorder to be started and stopped remotely. A continuous loop was mounted on the other recorder which could be controlled manually. The first track of this loop presented one of the training stimuli with an interstimulus interval of 500 msec. The second track produced a similar repetition of the other training stimulus. The choice between the two tracks was controlled manually.

The response apparatus consisted of a vertical panel with a response lever on each side of the front. In the case of a training with the /d-t/ stimuli, for example, the subject would find a picture of a roof ([dɑk] in Dutch) above one of the levers and one of a branch ([tɑk] in Dutch) above the other. In order to avoid response bias, these pictures were regularly switched during each training session.

In the case of the training with stimuli from the /b-p/ series the two pictures were, respectively, of a bin ([bɑk] in Dutch) and a parcel ([pɑk] in Dutch). Between each lever and corresponding picture were two bulbs, one red and one green. These served to give immediate knowledge of result at each response. Simultaneous with the pushing of a lever which was associated with the erroneous picture, a red flickering light began to burn. A green light burned continuously, simulutaneous with the pushing of a lever that was associated with the correct picture. A switch on the back of the panel could be manually operated by the experimenter in such a way that the subject could not see it. At each stimulus presentation and dependent on the positions of the pictures, it could be determined via this switch which of the levers corresponded to the correct picture.The experimenter's switch caused an electric circuit to come into operation by which relevant immediate knowledge of results in the form of a red or a green light could be realized. The experimenter could position the switch before the subject could answer, in view of the fact that his reaction time was shorter than that of the subject because of his perfect categorizing capacity. Secondly, the experimenter was supplied with a control list which gave the stimulus order of each random block.

The training sessions were hold in a sound quiet room.

4.3.3.3. Procedure.

The experimenter and the subject sat opposite one another at a table upon which the above-described response panel was placed. The subject sat facing the panels' front, while the experimenter faced its back. The subject could not see the manner in which the experimenter manipulated the apparatus.

Each subject had become accustomed to extensive listening tests as a result of the collection of perceptual data before the training.The stimuli were presented at the level of amplification that the subject found most comfortable. By means of the pictures and the corresponding contrasting training stimuli from the loop as well as stimulus support from the experimenter who emphasized the same contrasts loudly and clearly, it was made clear to the subject what was required of him/her, namely the learning to distinguish between the words [bɑk] and [pɑk] or [dɑk] and [tɑk]. The manner in which the response apparatus worked as well as the consequences of a response were explained to the subject. The first session than began. Between the collection of pretraining data and the first training session there was an interval of minimally one and maximally three months. Training was given daily with the exception of weekends. In each session two random blocks with a total of 40 stimuli were presented. The blocks were presented in random order. Immediately after a correct response, the program was continued with the following stimulus. After an incorrect response, the loop presented the subject with the falsely answered stimulus repeated four times while the experimenter pointed to the correct picture. In the case of regularly repeated errors, the loop repeatedly presented the subject with the two contrasting training stimuli in alternating order during which the experimenter pointed to the appropriate pictures. After one more repetition, the program was continued with the following stimulus (see Costello, 1977).

The experimenter noted every correct and incorrect response of a subject in a given session in a [+voice] and [-voice] stimulus-response confusion matrix.

After each session, the duration of which depended on the subjects' responses, the subject was rewarded with a candy. Although variable in length, an upper bound of 25 minutes was set for each session. As it had become apparent from preliminary research that hard-of-hearing children yield substantically better results during intensive listening tests given in the morning as opposed to those given in the afternoon, all of the experiments in this research took place in the morning.

4.3.3.4. Criterion.

For each subject the results of each session were subjected to a nonparametric calculation which produced a single index number for the confusion matrix of the two alternative forced choices. The index was based upon the ratio of correct to incorrect responses. This Pollack-Norman Index A' (Pollack & Norman, 1964), which lends itself well to the case of two-alternative-forced-choice experiments gives the extent to which one can detect a signal or stimulus property and is an ordinal approximation to the d-prime, the index of signal detectability from the theory of signal detection (Green & Swets, 1966).

The index varies from 0 to 1.0, with .50 indicating chance performance. The index is derived by considering both correct (hits) and incorrect (false alarms) acceptances of a unit (e.g. distinctive feature). For a two-alternative-forced-choice experiment the index for [+voice] and [-voice] performances is equal. The Pollack-Norman Index A' reflects, consequently, all signals to which the listener imputes common characteristics, whether he does so correctly or incorrectly.

The index is calculated from single hit and false alarm rates. The metric defines the hit to false alarm proportions of equivalent performance. In this way, the learning achievements of the experimental subjects, independent of individual strategies, can be compared. An equivalent performance of A'= .79 is, for example, obtained in the event of a hit to false alarm ratio of .80/.40, but also for .60/.20.

The index of performance will increase if the listener either increases his hit rate, decreases his false alarm rate, or both (see also Schultz & Kraat, 1973).

As a criterion for the discontinuation of the training it was decided that the maximum possible score of A'= 1.0 ought to be achieved in two consecutive sessions. Regardless of whether this criterion was met or not, training was suspended after the arbitrarily chosen number of 22 sessions.

A score of A'= 1.0 implies that a subject has made no errors in distinguishing between a [+voice] and a [-voice] stimulus during a given session. An A'= 1.0, consequently, indicates a 100% correct perception of the contrast between the two training stimuli. The distinction between acoustic and phonetic contrasts, however, is not expressed in this performance.

4.3.3.5. Results.

The results of this learning experiment are to be found in table 7.

subject	number of sessions to criterion	initial and final A'
1	12	.74 - 1.0
2	9	.79 - 1.0
3	7	.61 - 1.0
4	11	.69 - 1.0
5	20	.60 - 1.0
6	21	.88 - 1.0
7	22	.55 - .90

Table 7: Learning results of the four hard-of-hearing subjects (1-4) and the three speech impaired subjects (5-7)

One sees from these results that six of the seven achieved the criterion of perfect perception of a contrast between selected stimuli within the bounds of 22 sessions (subject 1 through 6)

The seventh experimental subject achieved a substantial improvement in contrast perception, but a perfect A'-score of 1.0 was not reached within the prescribed number of 22 sessions.

Moreover, it should be noted from table 7, that the most severely hard-of-hearing subjects (3 and 4) did not require the most sessions in order to reach the criterium level of 100% contrast perception. Table 7 demonstrates, furthermore, that the four hard-of-hearing children achieved the criterion level neither extremely quickly nor slowly,when an upper bound of 22 sessions was set (two-tailed binomial $p \geqslant .14$). In terms of the criterion, the three subjects whose speech was impaired were extremely slow in reaching the desired level ($p < .01$). It should be noted, however, that two of the three dysphasic subjects (5 and 7) started the training with a significantly worse initial level than the hard-of-hearing children. This can be concluded from their chance response performances during the first session. The remaining subjects, on the other hand, began the training with a better perceptual level and a response performance which was substantially better than pseudo-random answering ($p \leqslant .05$). These conclusions can be summarized as follows. In terms of the fixed criterion and the stimuli used, the selected hard-of-hearing children belong to the group of subjects who learn perceptual contrast moderately quickly. These children profit by a considerable perceptual foreknowledge. The extent of the disability appeared to have played a direct rôle in neither the learning tempo nor the perceptual state at the beginning of the training.

Considering the criterion and VOT stimuli, the experimental dysphasic subjects were, on the other hand, extremely slow in learning perceptual contrast. These children began the training, moreover, with a trial-and-error response behavior. The only dysphasic child with a favorable perceptual apriori-knowledge at trainings initiation (subject 6) required, nevertheless, a long-lasting training period of 21 sessions in order to reach the criterion.

Although the number of subjects here is far too small to substantiate any major conclusions, these quantitative differences between the two groups of auditorily handicapped children do suggest a qualitative difference in the processing of VOT stimuli. This difference is manifested in the higher initial level of performance and the shorter training time required to learn the distinction between temporally encoded speech stimuli by subjects with a defective transmission system compared with subjects with a disturbance in the specific speech processing mechanism.

One essential point in this research is the fact that, despite these differences, all subjects displayed a significant improvement in the perception of the contrast between the selected stimuli. The effect of this improvement will be considered in chapter 5.

CHAPTER 5

THE EFFECT OF THE TRAINING

This chapter deals with two aspects of the effect of auditory training that correspond with hypotheses 1 and 2 of chapter 2.4.

As regards hypothesis 1, it is claimed that in many cases the effect of specific auditory training, with the aid of stimuli of a VOT series, will improve perception of the voicing contrast. This improvement will first be manifest in the deviating identification functions, established before the training, of the /b-p/ and /d-t/ series that were composed by means of the VOT variables.

In chapter 5.1. the changes of these identification functions will be analyzed in the light of the established phoneme boundary criteria. This analysis will be carried out according to a before-after design.

Hypothesis 2 claims that, as a result of training with specific stimuli from a VOT series, the improvement of the perception of the voicing contrast is not restricted to this series. A transfer of acquired perceptual information is postulated, even if the stop consonants in question with their voicing characteristics occur in different vowel contexts. Moreover, a transfer is expected to word-initial stop consonants in different vowel contexts that are produced in another place of articulation than the trained stimuli. Hypothesis 2 is proved by means of some feature analyses. The data for these analyses have been obtained from a /b-p/ and a /d-t/ minimal pair list; the initial stop consonants of the words of these lists are part of normally spoken words and occur in different vowel contexts. The differences between the two members of these minimal pairs only consist in the voicing features of the initial stop consonants. Furthermore, a third list containing /b-p-d-t/ contrasts is used. These contrasts enable the listener to choose from a combination of the competing [+voice] and [-voice] features and the place of articulation, in this case of alveolar versus bilabial.

The predicted generalization of learned voicing information, associated with specific stop consonants, to similar phonemes, occuring in another context and produced in another place of articulation, is considered linguistically relevant in this research.

5.1. THE EFFECT OF TRAINING ON CATEGORICAL PERCEPTION

5.1.1. Collecting data.

The identification functions and discrimination curves of the /b-p/ and /d-t/ VOT series with the seven persons of the experimental group were determined before their training. This procedure has been described in chapter 4. By means of linear interpolation the identification functions produced four phoneme boundary values for each bilabial or alveolar contrast. These values (defined in chapter 2.4.) are the lower limit, the upper limit, the 50% cross-over point or the phoneme boundary, and the phoneme boundary zone, respectively.

The pre-training identification functions and discrimination curves are given in figure 7; the corresponding phoneme boundary values in table 8. The latter also mentions the results of the four listeners of the control group and the established criteria of the phoneme boundary values.

Collecting the post-training data was begun at least one day and not more than two days after the training had been finished. Post-training phoneme boundary values were determined in this way: if identification of a given series before the training produced a perceptual function for a subject which was normal as compared with the reference values used as criteria, this series was not presented again after the training. The reason was that hearing disordered children should not be overburdened any longer than is necessary by being subjected to monotonous listening experiments. With subject 1, 2 and 3, only the /d-t/ series was repeated; with subject 6 only the the /b-p/ series; and with subjects 4, 5 and 7, both series were repeated. The procedures were exactly the same as it had been when pre-training results had been obtained; they have already been described in chapters 3.3. for normally hearing listeners and 4 for experimental subjects.

The identification stimuli and two-step comparisons concerned were presented until, for each subject, 40 identifications were obtained per stimulus, and 32 two-step discriminations per comparison.

5.1.2. Results.

The post-training identification functions and the discrimination curves, and their relevant phoneme boundary values are given in figure 7 and table 8 respectively. The corresponding pre-training results are also given for immediate comparison.

Examination of the identification functions before and after training proves that in the case of subjects 1 and 2 there has been a realignment of the phoneme boundary of the /d-t/ series. An enhancement effect, and in some cases a sharpening effect, was produced in the identification functions of the two different VOT series with subjects 4, 5 and 7; the same can be noticed in the single graphs of subjects 3 and 6. Furthermore, it is seen that in most cases the discrimination curves after training better correspond with the relevant identification functions.

All these results point towards what has been predicted in hypothesis 1. At first sight the perception of the voicing contrast appears to have been corrected more or less in all experimental subjects. Furthermore, these preliminary results somehow imply the existence of a positive transfer of improved perception of the voicing contrast to stop consonants produced with a different place of articulation; this is because subjects 4, 5 and 7 show an enhancement effect on both VOT series, although they were trained on one VOT contrast only.

These results will be analyzed further in the next section.

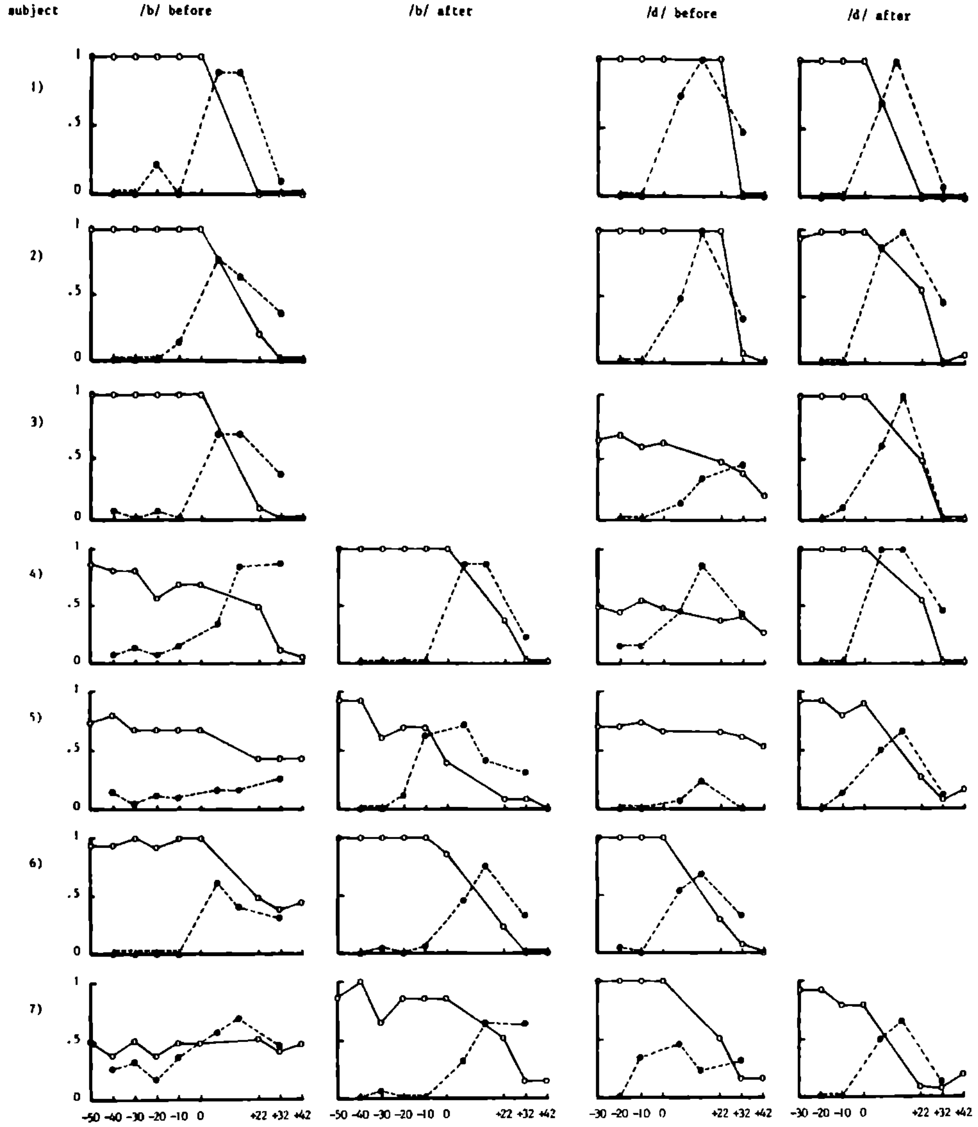

Figure 7: /b/ and /d/ identification functions (solid lines) and two-step discrimination curves (dotted lines) of four hearing disordered (1 - 4) and three speech disordered (5 - 7) subjects before and after discrimination training. Post-training scores have only been collected if the phoneme boundary parameters deviated from the criterion before the training (cf. table 6).
The discrimination curves are based on a comparison between two VOT stimuli that are two steps apart on the continuum. The dots of these curves are shown halfway this distance.

			/b-p/series				/d-t/series			
			75%	50%	25%	zone	75%	50%	25%	zone
	criterion		-7,00/+10,50	-1,16/+17,80	+8,28/+24,50	---/21,05	-3,88/+11,10	+2,54/+18,30	+9,24/+26,38	---/19,56
Experimental Group of hearing-disordered children	subj 1	a	+ 5,50	+11,00	+16,50	11,00	+24,50	+27,00	+29,50	5,00
		b					+ 5,50	+11,00	+16,50	11,00
	2	a	+ 6,88	+13,75	+20,63	13,75	+24,78	+27,56	+30,33	5,55
		b					+13,75	+23,67	+27,83	14,08
	3	a	+ 6,11	+12,22	+18,33	12,22	---	+22,00	---	---
		b					+11,00	+22,00	+27,00	16,00
	4	a	-28,00	+20,00	+28,25	56,25	---	---	---	---
		b	+ 9,17	+18,33	+25,75	16,58	+13,75	+23,67	+27,83	14,08
Experimental Group of speech-disordered children	5	a	---	+22,00	---	---	---	---	---	---
		b	-35,00	- 3,33	+11,00	46,00	+ 5,50	+14,67	+24,50	19,00
	6	a	+10,00	+20,00	---	---	+ 7,86	+15,71	+24,50	16,64
		b	+ 5,08	+13,54	+22,00	16,92				
	7	a	---	---	---	---	+11,00	+22,00	+30,33	19,33
		b	+ 3,67	+22,00	+29,58	25,91	+ 1,57	+ 9,43	+17,29	15,72
Control Group of hearing-disordered children	8		+ 6,11	+12,22	+18,33	12,22	+18,33	+24,86	+28,43	10,10
	9		+ 6,11	+12,22	+18,33	12,22	+18,33	+24,86	+28,43	10,10
	10		+ 5,50	+11,00	+16,50	11,00	+23,67	+26,44	+29,22	5,55
	11		-13,75	- 8,33	- 4,17	9,58	+24,50	+27,00	+29,50	5,00

Table 8: Results before (a) and after training (b) of the phoneme boundary parameters of the /b/ and /d/ identification functions of four hearing disordered and three speech disordered subjects of the experimental group and the boundary parameters of four hearing disordered subjects of the control group.
Post-training data were collected only if the results before training deviated from the criterion. At the top of the table the ranges of the allowed values of table 2 are given by way of reference per boundary value; - - - means that this point is lacking in the identification function.
Subjects 1 - 5 were trained on /d-t/ stimuli; subjects 6 and 7 on /b-p/ stimuli.

5.1.3. Analysis of the results.

The results of the changes of the perceptual functions were analyzed twice.
In the first analysis the changes of phoneme boundary values (as mentioned in table 8) were examined in their relation to the established criteria.
In the second analysis the differences were expressed in terms of changes of the entire perceptual function, by means of Categorical Score C (Fourcin et al., 1978; Simon & Fourcin, 1978; cf. also chapter 4.2.).
For the changes of phoneme boundary values before and after training a non-parametric test for related samples was used, i.e. the McNemar-test for the significance of changes (Siegel, 1956). This test is applicable especially in those 'before and after' designs in which each person is used as his own control.
It did not seem realistic to expect that all deviating boundary values would be normal after training. Therefore the application of the McNemar-test was based on two different criteria. In the first place only those changes of phoneme boundary values were

included in the analysis which were outside the scope of the VOT criterion before the training but inside the scope after the training (strict criterion). Secondly, all positive changes after the training were considered, whether inside the scope of the VOT criterion or approaching these (approximation of criterion).

28 observed data of table 8 can be placed in a McNemar table of frequencies, the dependent variable being the four boundary values of the seven series in which training had been carried out. The four possible before-after situations are: + + (positive before and after training, i.e. no change); - - (negative before and after training, i.e. no change); + - (positive before training, negative after training, i.e. a negative change); - + (negative before training, positive after training, i.e. a positive change). For statistic analysis only the change scores are used.

Analysis of significance of changes was carried out on the data of the trained series; in the same way the phoneme boundary values of the three untrained series were analyzed. The latter series differed from the former ones as regards place of articulation, and they were also presented after the training. These three series supplied 12 observations that can be interpreted within the framework of a possible transfer. The significance of changes for the boundary values of the three untrained series were also examined, both on strict criterion and on approximation of criterion. Table 9 shows the frequencies observed, as well as the results of the McNemar-test for the data of the trained and untrained series, based on the two different criteria. In all cases there is a predicted tendency of the changes in accordance with hypothesis 1 and 2. Therefore, a one-tailed test was carried out.

	before-after changes	+ +	- -	+ -	- +	McNemar-test
trained	change to criterion	3	11	0	14	χ^2=12,07 df=1 one tailed $p < .0005$
	approximation of criterion	3	1	0	24	χ^2=22,04 df=1 one tailed $p < .0005$
non trained	change to criterion	2	5	0	5	one tailed $p = .03^*$
	approximation of criterion	2	0	0	10	χ^2= 8,10 df=1 one tailed $p < .005$

Table 9: McNemar-test for the significance of changes of before-and-after training results according to strict criterion and according to approximation of criterion. The changes of the four boundary values of 7 trained VOT series have been examined. Also the positive transfer to the results of 3 untrained VOT series has been examined. The latter series differed from the trained series in place of articulation (cf. table 6).
The asterisked p-value has been calculated with the binomial-test because of the low frequency of the changes.

From table 9 it appears that the positive changes of the phoneme boundary values before and after training are significant both according to strict criterion and according to approximation of criterion. This is the case both for the trained and for the non-trained series. Therefore this analysis gives strongly foundation to hypothesis 2, which is actually studied only later, during the examination of a linguistically relevant effect of training.

The significant changes of the phoneme boundary values of table 9 justify acceptance of hypothesis 1, which claims that specific training with stimuli of the two VOT series realizes an improvement of the categorical perception of these series. Table 9 also implies that there has not been any negative change in the boundary values. Generally speaking, therefore, selective training appears to have a significantly positive, dynamic effect on the boundary values.

For the second analysis Categorical Score C was used (cf. chapter 4.3). For each subject the percentage of correct identifications of the relevant /b-p/ and /d-t/ series was determined per stimulus, before and after the training. In accordance with the boundary criteria (determined on the basis of the identification functions of normally hearing subjects, cf. chapter 3.4) the last three VOT stimuli of the /b-p/ and /d-t/ series were classified as voiceless; the first six stimuli of the /b-p/ series and the first four stimuli of the /d-t/ series were classified as voiced.

subject		C-scores	
		/b-p/	/d-t/
1	a	1.00	.71
	b		1.00
2	a	.96	.69
	b		.79
3	a	.96	.27
	b		.86
4	a	.50	.10
	b	.91*	.83
5	a	.28	.14
	b	.56*	.69
6	a	.67	.89
	b	.92	
7	a	-.03	.74
	b	.59	.71*

Table 10: Categorical Scores C before (a) and after (b) training, in accordance with the perceptual functions of the 7 experimental subjects of table 8 and figure 7. The three asterisked C-scores are based on transfer.

These references had a double effect: first, it was possible to establish the degree of Discrimination (D) of a perceptual function in terms of the average percentage of all correct identifications; secondly, with this D the categorical score C could be calculated per identification function and for each subject before and after training. These C-scores of seven experimental subjects are given in table 10.

A negative C refers to a reversal of the perceptual function (/d/ is perceived as /t/ or vice versa); a C = 0 to a chance response performance; a C = .50 to a sloping function; a C ⩾.75 to a categorical identification function with a sufficiently abrupt transition between two categories.

This classification of the C-score enables evaluation, in significant terms, of changes in the perceptual function before and after training, according to McNemar's before-and-after design (Siegel, 1956). However, even if this classification is used, it is not to be expected that every series, a priori perceived non-categorically, will be perceived categorically after training. Therefore, C-scores are analyzed, not only on the criterion of 'change into a categorical mode' but also on the criterion of 'general improvement'. In the latter case all C-scores that occur in a higher classification after training are considered positive changes (i.e. - +).

In accordance with these two criteria the results of the seven trained series and the three non-trained ones of table 10 are distributed among the relevant cells of table 11.

	before-after changes	+ +	- -	+ -	- +	McNemar-test
with training	change into a categorical mode	0	2	0	5	one tailed p= .03
	general improvement	0	0	0	7	one tailed p= .008
without training	change into a categorical mode	0	2	0	1	insufficient n for statistical test
	general improvement	0	2	0	1	insufficient n for statistical test

Table 11: Effect of training expressed in changed Categorical Score of 7 listeners before and after training and transfer results in C-scores of three listeners. With $n < 10$ changes, p is based on the binomial test.

Examination of the cell frequencies in table 11 indicates that after training five out of seven identification functions changed positively from a non-categorical to a categorical mode of perception. This change is significant (p= .03).

From table 10 it appears that the greatest change occurred in severely hearing-disordered subjects 3 and 4, who jumped from a sloping or chance response level (C=.27 and C=.10 respectively) to a categorical level (C=.86 and C=.83 respectively).

Subjects 5 and 7 did not reach a categorical level according to the strict criterion and remained in the (- -) cell, therefore; nevertheless, they show considerable difference in identifying the relevant series before and after training. The respons level of subject 5 changes from chance response (C=.14) to approximation of the categorical level (C=.69), whereas that of subject 7 changes from chance response (C= -.03) to a sloping response level (C=.59).

It is evident, therefore, that according to the criterion of general improvement, table 11 shows a positive, significant change of all seven perceptual functions (p= .008).

In accordance with the first analysis the results of the second analysis give sufficient evidence to accept hypothesis 1. The results of the three non-trained series are not suitable for statistical analysis because of the low number of observations. In this case descriptive analysis is the only possibility of explanation.

In subject 4 the identification function of the non-trained series changes from a sloping mode (C=.50) to a categorical one (C=.91). This points to a transfer of learned information. Subject 5 shows a considerable numerical jump from C=.28 to C=.56 but remains within the class that corresponds with a sloping identification function. Subject 7 shows a slight decline of C=.74 to C=.71 and thus keeps approaching the C-score of a significant categorization very closely but remains within the class of a sloping identification function.

If both criteria are maintained it appears that, on the basis of the C-scores of the three non-trained series:

a) in one case there is a change from a non-categorical to a categorical mode of perception.

b) in another case there is a numerical though not a significant trend to improvement of perception.

c) in the last case there is a permanent, relatively close approximation of the criterion of categorical perception.

These C-score results do not justify to conclude generally that there has been an evident form of transfer of information to the non-trained series. However, the observed positive trends in the perception of these series in the first and second analysis point that way. Besides,there has not been any substantial decline in labeling results.

It appears from the post-training perceptual functions (cf. table 10) that, in spite of evident improvement, two out of three speech-disordered children did not reach a categorical level for the trained and non-trained series of stimuli after training, but all four hearing-disordered children did. This suggests that the effect of training of time-coded speech stimuli with dysphatic children was different from that of hearing-disordered children. The same difference had already been suggested by the learning results in chapter 4.3.3.5.

5.2. THE LINGUISTICALLY RELEVANT EFFECT OF THE TRAINING.

5.2.1. Collecting the pre-training and post-training data.

The data for indicating a linguistically relevant effect of the training were obtained by means of three different lists of minimal pairs.

The first list included 26 two-alternative forced choice items with semantic meanings. Each pair of words consisted of the same phonemes, but the voicing features of the word-initial prevocal stop consonant were different (e.g. /dɔr - tɔr/). The initial stop consonants of the first list always consisted of the homorganic alveolar stops /d/ and /t/. The vowels after the initial stops varied for each pair. The minimal pairs consisted of words of one or more syllables. The phonetic representation of the 26 minimal pairs is given in Appendix A; the words have been transcribed according to the International Phonetic Alphabet.

The subjects were given printed lists of the minimal pair alternatives. The 26 items had been arranged at random in /d-t/ or /t-d/ pairs.

A tape was constructed in addition to the first list. On this tape one of the two possible words of each of the 26 alternative choices had been spoken in isolated position. This tape comprised 12 words beginning with a [+voice] alveolar stop consonant, and 14 words beginning with a [-voice] one. The interval between the stimuli lasted four seconds. The succession of [+voice] and [-voice] stimuli was randomized. The words were spoken by the same male speaker who had produced the original VOT stimuli of the /b-p/ and /d-t/ series and the VOT training stimuli of this research. The recordings were made with a AKG-CK1 condensator microphone and a Studer B62 tape-recorder. Just as in the other listening sessions in this research, the auditory stimuli were presented individually by means of a Revox A-77 tape recorder and a set of AKG K150 headphones.

For each subject, sound volume had been adjusted beforehand at its most comfortable level.

The subjects were instructed what to do. After the presentation of each word they were to mark on the list the relevant word of the alternative choice that they thought they had heard. Speed of presentation of the stimuli could be adapted to the speed of each subject through remote control of the tape recorder by means of a pedal.

The second list comprised 42 minimal pairs with semantic meanings. The words consisted of one or more syllables. The word-initial prevocal stop consonants of these words were the bilabials /b/ (voiced) and /p/ (voiceless). On the tape used for this list 21 words began with a [+voice] , and 21 with a [-voice] stop consonant. Conditions and procedure were the same as with the list of alveolars. Appendix B gives the transcribed International Phonetic Alphabet representations of the 42 /b-p/ pairs.

The third list comprised 16 items of four semantically meaningful words of one or more syllables. Every item of four words consisted of the same phonemes, except for the initial prevocal stop consonants. These consonants were the phonemes /b/, /p/, /d/ and /t/. They were followed by different vowel contexts, but the words of each four-alternative item had the same vowel context (cf. Appendix C). The phonemes contrasted per item as regards voicing (e.g. /b-p/), place of articulation (e.g. /b-d/), or voicing and place of articulation jointly (e.g. /b-t/).

There were four different copies of the printed list of four-alternative items. In each list and for each item the sequence of the words beginning with /b/, /p/, /d/ or /t/ had been arranged in ramdom order. To each of these four lists a separate tape was added with auditory stimuli that consisted of one of the four possible alternatives per item. The tapes had been made in such a way that after a complete presentation of the four lists each of the four alternative words per item had been presented once. After the presentation of a stimulus word the subject pointed to the word of the four alternatives that he thought he had heard. The responses of the subjects were noted down directly in a joint hit and false alarm confusion matrix.

The pre-training data of the /b-p/ and /d-t/ lists for all seven experimental subjects and the four control subjects were collected in the same period as the other pre-training data (cf. chapter 4). The pre-training data of the /b-p-d-t/ lists were collected only for the experimental subjects in the same period. The presentation of the six respective /b-p/, /d-t/ and /b-p-d-t/ lists took place per subject in random order and was completed in four days. The /d-t/ and /b-p/ lists were presented four times before the training; so, per subject the results consisted of 48 responses on [+voice] presentations and 56 responses on [-voice] presentations for the /d-t/ list; for the /b-p/ list they consisted of 84 responses on [+voice] presentations and 84 responses on [-voice] presentations.

Each of the four /b-p-d-t/ lists was presented four times to each experimental subject.

Analysis of the results of these lists was only made on the performances of those experimental subjects who appeared to have improved significantly on the /b-p/ and /d-t/ lists after the training. This appeared to be the case with six out of seven experimental subjects (cf. chapter 5.2.2.1). Eventually these six subjects together provided 384 responses (correct and incorrect ones) for each of the /b/, /p/, /d/ or /t/ stimuli.

The post-training data of the /d-t/, /b-p/ and /b-p-d-t/ lists of the experimental subjects were collected within four days and in the same period as the data of the post-training VOT series: so, collecting the post-training data was started at least one day but not later than two days after the training had been finished, and at least one month but not later than three months after the pre-training data had been collected.

As regards the four control subjects who had not had specific auditory training, a two months' latency period was introduced between the collection of the first and second set of data obtained from the /b-p/ and /d-t/ lists.

The post-training and post-latency period data for experimental and control subjects respectively were collected in exactly the same way as the pre-training ones, until it was possible to dispose of the same number of data as had been obtained before the training.

5.2.2. Feature analysis of the results with the /d-t/ and /b-p/ lists.

According to a before-after design two different feature analyses were made on the results obtained with the /b-p/ and /d-t/ lists.

First, the significance of the difference between changes was calculated by comparing the average gain-score, calculated for the whole experimental group, with the average gain-score of the whole control group. In this calculation the combined results of the [+voice] and [-voice] identifications, respectively, of the /d-t/ and /b-p/ lists were compared separately and independent of the place of articulation.

Secondly, the gain-score of the [+voice] and [-voice] performances of each experimental and control subject were determined separately, and subjected to a criterion that is considered linguistically relevant in this research.

5.2.2.1. Analysis of the gain-scores.

Guilford adduced arguments (1965) to test the significance of the difference between changes with the help of gain-scores in an experiment with non-matched experimental and control subjects. He preferred this method to e.g. the calculation of the significance between the final test performances of the experimental group (henseforth to be called E-group) and the control group (to be called C-group).

One of the important advantages is that the initial differences between the E-group and the C-group are taken into account when the average gain-score is calculated. Although this is not essential for the final conclusions of the first analysis, which are based on the observations of the gain-scores, yet we shall mention the differences between the initial and final performances within the E-group and C-group, respectively, and those between the E-group and C-group before and after the training (cf. tables 12 and 13).

For a before-after comparison within the E-group and C-group respectively, a related sample-design was used, viz. a paired t-test based on the mean of the item differences before and after training of the subjects in each respective group (for this test, cf. chapter 3.4.2).

It had been predicted that the E-group would achieve a better performance in perceiving the voicing contrast after training than before; for that reason a one-tailed test was used. There was some uncertainty about the results before and after training of the C-group, however. Nothing had been predicted about the difference in performance, but it is possible that frequent presentation itself of the perceptual contrast improves discrimination of it. A similar thing could be observed in the exposure phenomenon, cf. chapter 4.2. (cf. also Ades' theorie, 1977). Therefore both a one-tailed and a two-tailed test is performed on the results of the C-group.

A t-test for independent samples, based on group averages (Silverman, 1977) was used for the comparison between the results of the E-group and C-group before the training, and between those of the two groups after training.

When the two groups are compared before training it is not predicted which of them will be superior; actually, a large difference between the two groups is even undesirable. In this case a two-tailed test is used. When the results of the E-group and the C-group are compared after training and after the latency period respectively, it is predicted that the results of the E-group will be better; in this case a one-tailed test is used. The results of the E-group and the C-group of the correctly identified [+voice] presentations of the /d-t/ and /b-p/ lists are given in table 12.

	Correct [+voice] responses of the Experimental Group		Correct [+voice] responses of the Control Group		Difference ($M_E - M_C$)	t	df	p
	Mean	SD	Mean	SD				
Before	95,14	30,85	107,75	5,68	-12,61	-.79	9	>.05(two tailed)
After	114,71	22,06	114,25	16,03	0,46	.04	9	>.05(one tailed)
Gain($\bar{D}$)	19,57	21,72	6,50	15,55	13,07	1.05	9	>.05(one tailed)
$t_{\bar{D}}$	2.38		.84					
df	6		3					
p	<.05(one tailed)		>.05(one tailed) >.05(two tailed)					

Table 12: Test of significance of the difference between gain-scores of all correctly perceived voiced contrasts, independent of their place of articulation. In the Experimental Group there are seven combined [+voice] observations; in the Control Group there are four. The absolute maximum of correct [+voice] responses is N = 132.

Besides the changes within the same group and between different groups which are to be seen in this table, the gain-scores of the E-group (19,75) and the C-group (6,50) show an mean difference of 13,07 scores in favour of the E-group. This difference is not significant. The final, important conclusion of the group analysis about the correct [+voice] identifications with the /d-t/ and /b-p/ lists is that the E-group shows a larger gain of [+voice] identifications after training than the non-trained C-group; this gain is not significantly different from that of the C-group, however.
The combined results of the correct [-voice] identifications obtained from the /d-t/ and /b-p/ lists for the E-group and C-group are given in table 13.

	Correct [-voice] responses of the Experimental Group		Correct [-voice] responses of the Control Group		Difference (M_E - M_C)	t	df	p
	Mean	SD	Mean	SD				
Before	72,29	23,17	80,00	26,73	-0,71	-.05	9	>.05(two tailed)
After	113,57	24,63	90,25	14,61	23,32	1.77	9	>.05(one tailed)
Gain($\bar{D}$)	34,29	10,66	10,25	21,79	24,04	2.51	9	<.05(one tailed)
$t_{\bar{D}}$	-8.51		-.94					
df	6		3					
p	<.01(one tailed)		>.05(one tailed) >.05(two tailed)					

Table 13: Test of significance of the difference between gain-scores of all correctly perceived voiceless consonants independent of their place of articulation. There are seven combined [-voice] observations in the Experimental Group and four in the Control Group. The absolute maximum of correct [-voice] responses is N= 140.

Comparison between the gain-scores of the E-group (34,29) and the C-group (10,25) shows a gain for the E-group that exceeds that of the C-group significantly with 24.04 scores.
The joint findings of the [+voice] and [-voice] feature analyses of the whole E-group and the C-group lead to the conclusion that the E-group profited more after training in identifying voiced and voiceless items than the non-trained C-group.
The observed changes in the hit and false-alarm ratios of the group responses clarify this conclusion. On account of the [+voice] identifications the hit-rate of the E-group appears have increased from 72% to 87%; however, this is not significantly different from the change in the hit-rate from 82% to 87% in the C-group. On the other hand,the false-alarm rate in the E-group appears to have decreased after training from 48% to 19%, which is significantly different from the decrease of the false-alarm rate in the C-group from 43% to 36%.
The determination by means of these data of the Pollack-Norman Indices A' as a degree of signal detectability (Pollack & Norman, 1964; cf. chapter 4.3.3.5. for explanation)

points to an improved index of performance from .70 to .91 in the E-group i.e. to a gain of A'= .21. In the C-group the index of performance changes from .79 to .85; this means a gain of A'= .06.
Because of the two-alternatives forced choice task, the Pollack-Norman Index of the [+voice] and [-voice] performances is equivalent; therefore, the same A' changes are valid both for the [-voice] identifications and for the [+voice] performances described.

The data for the feature analysis of the results of the seven individual experimental subjects and the four individual control subjects are given in table 14. These data have been arranged in such a way that, for the experimental subjects, the first six columns are based on before-after results of the [+voice] and [-voice] identifications that corresponded, as regards place of articulation, with the respective stimuli on which each of them had been trained. This means that the data of each of the first five experimental subjects are based on 48 initial /d/ and 56 initial /t/ presentations in a varied vowel-context. They have been obtained from the list of /d-t/ contrasts that had been presented four times before and after training. The data of experimental subjects 6 and 7 are based on 84 /b/ and 84 /p/ presentations in a varied vowel context. They too have been obtained from the /b-p/ list that had been presented four times.
It is exactly the other way round for the experimental subjects referred to in the last six columns of table 14. These data are based on before-after results of the [+voice] and [-voice] identifications which differed from the training stimuli in place of articulation.
No individually performed arrangement was required for the control subjects. Their data before and after two months'latency period are separate as regards their responses to the lists of /d-t/ and /b-p/ contrasts that had been presented four times.
Table 14 gives, for each experimental and control subject the per cent of correct identifications per voicing feature of the /d-t/ and /b-p/ lists before and after training, and before and after the latency period, respectively.
A separate statistic calculation of the significance of changes was performed for each subject on the [+voice] and [-voice] results of each list. In this case a related sample design was chosen, and a paired t-test was performed on the item-to-item changes that were based on the four presentations before training and the four after training (Silverman, 1977). Since in this way all separate changes of the relevant items were involved in the test of significance, it could be assumed that a change of the ability to extract the voicing features in the varied vowel contexts could indeed be analyzed. Improvement of the perceptual extraction of the voicing features that form part of a consonant in varied vowel contexts was predicted in hypothesis 2.
The results of the analysis are given in table 14. It had been predicted in hypothesis 2 that the experimental subjects will produce more correct [+voice] and [-voice] identifications after training than before, irrespective of the vowel context and the place of articulation of the stimuli. For this reason a one-tailed test was applied

here. On the results of the control subjects a two-tailed test was performed, because no prediction in a specific direction has been made in this case.

The significance test of the changes of the separate items of every list was not the only criterion, however. A second criterion was introduced in combination with the first. On account of the second criterion it could be assumed that improved recognition of the [+voice] or [-voice] features was relevant or not for a subject.

	subjects	trained contrast	voice	% correct [+voice] and [-voice] identifications of consonants with the same place of articulation as the trained stimuli				nontrained contrast	voice	% correct [+voice] and [-voice] identifications of consonants with a different place of articulation as the trained stimuli			
				before-after	$t_{\bar{D}}$	df	one tailed p			before-after	$t_{\bar{D}}$	df	one tailed p
EXPERIMENTAL GROUP	1	t-d	+	92 - 97	1.20	11	>.05	p-b	+	76 - 90	1.81	20	<.05
			-	43 - 95	5.12	13	<.01		-	57 - 79	2.68	20	<.01
	2	t-d	+	100 - 100				p-b	+	93 - 100	1.83	20	<.05
			-	63 - 91	3.51	13	<.01		-	48 - 89	5.97	20	<.01
	3	t-d	+	75 - 50	-1.39	11	>.05	p-b	+	48 - 57	0.53	20	>.05
			-	21 - 71	3.32	13	<.01		-	57 - 52	-0.33	20	>.05
	4	t-d	+	25 - 83	2.88	11	<.01	p-b	+	29 - 68	2.38	20	<.01
			-	21 - 75	5.49	13	<.01		-	38 - 38	0.16	20	>.05
	5	t-d	+	92 - 100	1.00	11	> 05	p-b	+	81 - 90	1.45	20	>.05
			-	57 - 86	1.75	13	<.05		-	76 - 95	2.17	20	<.05
	6	p-b	+	86 - 95	1.00	20	>.05	t-d	+	92 - 100	1.00	11	>.05
			-	81 - 100	2.17	20	<.05		-	71 - 100	2.28	13	<.05
	7	p-b	+	76 - 95	1.92	20	<.05	t-d	+	58 - 98	2.76	11	<.01
			-	76 - 63	-1.19	20	>.05		-	64 - 79	1.07	13	>.05

	subjects	list	voice	% correct alveolar voicing identifications before and after a two-month interval without training				list	voice	% correct bilabial voicing identifications before and after a two-month interval without training			
				before-after	$t_{\bar{D}}$	df	two tailed p			before-after	$t_{\bar{D}}$	df	two tailed p
CONTROL GROUP	1	t-d	+	83 - 77	-0.82	11	> 05	p-b	+	86 - 73	-1.43	20	>.05
			-	50 - 64	0.79	13	> 05		-	33 - 46	1.07	20	>.05
	2	t-d	+	75 - 98	1.69	11	> 05	p-b	+	90 - 98	1.37	20	>.05
			-	57 - 71	1.00	13	>.05		-	33 - 55	2.09	20	<.05
	3	t-d	+	67 - 83	1.00	11	>.05	p-b	+	81 - 75	-0.71	20	>.05
			-	79 - 80	0.17	13	>.05		-	57 - 77	2.09	20	<.05
	4	t-d	+	90 - 96	1.00	11	>.05	p-b	+	76 - 96	2.12	20	<.05
			-	93 - 59	-2.72	13	<.05*		-	71 - 68	-0.28	20	>.05

Table 14: *Above*: the percentage of correct [+voice] and [-voice] identifications of experimental subjects on /t-d/ and /b-p/ lists before and after training of the voicing contrast. Also the results of the test of significance of the changes, calculated for the separate items of each list are given.
Below: the results of the control subjects on the same lists, before and after training. The asterisk indicates a significant difference that is based on a higher score before the latency period.

The second criterion was determined by means of the 5% confidence limits of the total possible number of correct [+voice] or [-voice] responses, based on the four presentations of the /d-t/ or the /b-p/ list. This criterion was applied in combination with specific results of the paired t-test on the mean changes of the item scores in table 14.

The following rules of thumb were used for the combined criterion. A performance by a subject was considered negative if the corresponding score did not exceed the established upper limit of the confidence interval. This confidence interval was calculated with the formula of the z-approximation of the binomial distribution,which had been corrected for continuity,with p= .05 and with n observations per feature (Siegel, 1956). N depended on the number of items in each list. In practice such a negative performance in identifying two features of a contrast means either an ambivalent chance performance or a systematic substitution of the presented feature by a contrasting one. In such a case the presented feature is not correctly perceived in a categorical mode. A specific performance was considered positive and perceived categorically correctly if the corresponding score exceeded the established upper limit of the confidence interval.

Consequently, a change was judged positive if a performance was negative before training, and positive after training because the number of correct identifications had crossed the upper limit of confidence. A negative change could take place in the opposite direction. This kind of positive or negative change is called a categorical jump.

The significant results of the paired t-test were considered relevant only if they corresponded with a positive or negative categorical jump according to the binomial-criterion. Moreover, those t-test results were considered relevant which reflected a significant positive or negative change without a categorical jump within the region of a positive performance that had been determined according to the binomial-criterion. In this case a positive or negative change could be considered approximation of or removal from a perfect, stable perception of the relevant feature.

Significant changes resulting from the paired t-test, which were realized within the region of a negative performance, were not involved in the analysis on account of problems concerning their interpretation. They were considered irrelevant.

If a positive or negative categorical jump did not correspond with a significant t-test, this change was not involved in the analysis.

To determine the limits of confidence with the binomial-test, a one-tailed test was performed on the results of the experimental subjects, in view of the predicted tendency of their results. A two-tailed test was performed on the results of the control subjects. An experimental subject appeared to have achieved a positive [+voice] performance on the /d-t/ list if he correctly identified at least 63% of the [+voice] presentations, and a positive [-voice] performance if he did so with at least 62% of the [-voice] presentations. As to the /b-p/ list, it was considered a positive [+voice] or [-voice] performance if at least 60% of the identifications was correct.

For a positive performance the control subjects had to give at least 65% correct identifications for [+voice] presentations on the /d-t/ list, and at least 64% correct identifications for [-voice] ones; with the /b-p/ list, the amount had to be at least 61% both for [+voice] and for [-voice] presentations.

On the base of the double criterion determined, the results of table 14 could be rearranged in a McNemar table, corresponding with the relevant changes that had been observed per subject.

	before-after changes	+ +	− −	+ −	− +	McNemar-test
Experimental Group	same place of articulation as trained contrast	5	0	0	8	p=.004 (one tailed)
	different place of articulation as trained contrast	3	3	0	8	p=.004 (one tailed)
Control Group	combined bilabial and alveolar results	9	2	1	2	insufficient changes for statistical test

Table 15: Analysis of significance of relevant changes for the combined [+voice] and [-voice] results of table 14. Separate analyses have been carried out, first on those stimuli of which the initial stop consonant corresponded with the training stimuli as regards place of articulation; secondly, on those stimuli of which the initial stop consonant differed in place of articulation from the training stimuli.
In the control group the results of bilabial and alveolar stop consonants have been combined.

For this arrangement, the results of the experimental subjects have been combined in the first part of table 15 for those [+voice] and [-voice] stimuli of which the initial stop consonant was realized in the same place of articulation as the training stimuli. This applied to thirteen out of fourteen observations, since the [+voice] results of subject 3 showed a negative categorical jump that was not in accordance with the double criterion.

On account of the combined criterion and the one-tailed McNemar test, the eight significantly positive changes after training appear to point to positive learning of the group of experimental subjects (p=.004); besides, there had been no significantly negative changes. The changes consist of five positive categorical jumps and three changes to perfection. These results of the seven individual experimental subjects conform what had been primarily claimed in hypothesis 2, viz. that perceptual learning of the voicing contrast which has been trained in homorganic initial stop consonants in a specific vowel context, generalizes to the same consonants in varied vowel contexts.

When the initial consonants of [+voice] and [-voice] stimuli differed from the training stimuli in place of articulation, the results were combined as well. In this way 14 observations for the McNemar test were supplied.

It appears from table 15 that the eight significantly positive changes can also be interpreted as positive learning by the group of experimental subjects (p=.004); besides, there had been no significantly negative changes. The changes concern four positive categorical jumps and four changes to perfection. This enables confirmation of the second thesis of hypothesis 2. This thesis claimed that in the case of auditory training with two homorganic initial stop consonants, a transfer would occur from learned voicing information to other homorganic stop consonants which differed from the training stimuli in place of articulation and in vowel context.

On further examination of the 16 significant positive changes occurring in the group of the experimental subjects, it appears that 10 concern an improvement of [-voice] identifications, and 6 concern an improvement of [+voice] identifications. Six out of eight performances of the experimental subjects which according to the criterion remained positive before and after training were associated with [+voice] stimuli; of three performances which remained negative, two were associated with [-voice] stimuli. So, in general, the problems before training were about recognizing the voiceless stimuli rather than the voiced ones. On the average, the number of incorrect identifications was 16% higher with the voiceless stimuli. This could already be expected when the pre-training identification curves of the bilabial and alveolar series in figure 7 were examined. It could also be expected on account of the analysis of the gain-scores which were based on the separate voicing features of the /d-t/ and /b-p/ lists in chapter 5.2.2.1.

Furthermore, the severely hearing-disordered subject, no. 3, did not show any satisfactory general perceptual learning. On the /t-d/ list a negative categorical jump from 75% to 50% is observed for the [+voice] identifications. This jump does not correspond with the combined criterion, however. The /t/ identifications show significant improvement, but the /p/ and /b/ identifications remain within a chance performance that is considered negative. On account of these dubious results of subject 3, her responses were eliminated when the /b-p-d-t/ list was analyzed.

Neither did the severely hearing-disordered subject 4 learn completely as expected. He learned to distinguish the voicing contrast in stop consonants that corresponded with the training stimuli as regards place of articulation. On the transfer list he showed a positive categorical jump in the [+voice] stimuli, but the result in the [-voice] stimuli remained negative.

The improved identification of one feature of a perceptual contrast after auditory training need not imply that the contrasting feature will also be identified better, as can be observed with subjects 3 and 4. This fact somehow supports the suggestion that the two voicing features are processed by a double detector mechanism (Eimas et al., 1973; Eimas, 1974; Cutting & Eimas, 1975).

The results in table 14 of the four subjects from the control group have also been arranged according to the combined criterion in the McNemar table (15). Two changes were eliminated according to this criterion, so that 14 observations remained for the analysis of changes. The arrangement shows that, even before the latency period, the control group scored nine times positively according to criterion.

The large amount of positive performances before and after the latency period, and the distribution of the other results among the remaining cells of table 15 make it impossible to interpret the changes of the results of the individual subjects in the control group by comparing them with the results of the subjects in the experimental group.

The final conclusion of the second analysis of the results of the /d-t/ and /b-p/ lists is that after auditory training six out of the seven experimental subjects have learnt to identify one or both of the voicing features of the voicing contrast in an initial stop consonant in a better and relevant way, irrespective of the vowel environment of this consonant or its place of articulation.

The combined twofold criterion does not enable a comparison of the results of the individual experimental subjects with those of the individual control subjects.

5.2.3. Analysis of the results with the /t-d-p-b/ lists.

In view of the questions dealt with in this section, the results of the /t-d-p-b/ lists of only those six experimental subjects were included for analysis who showed positive learning with the identifications of the /t-d/ and /b-p/ lists. For that reason, the responses of subject 3 were eliminated (cf. chapter 5.2.2.1).

On account of the combined responses of these six subjects, 384 identifications per phoneme were at our disposal both before and after training. The phoneme confusion matrices before and after training of these combined responses are given in table 16.

		response				N
		/t/	/d/	/p/	/b/	
stimulus	/t/	186	127	30	41	384
	/d/	49	257	27	51	384
	/p/	24	28	209	123	384
	/b/	21	39	37	287	384

BEFORE

		response				N
		/t/	/d/	/p/	/b/	
stimulus	/t/	280	47	34	23	384
	/d/	35	320	10	23	384
	/p/	28	10	282	64	384
	/b/	15	31	27	311	384

AFTER

Table 16: The phoneme-confusion matrices of the observed before- and after training results on the /t-d-p-b/ lists of six experimental subjects.

A feature analysis was first performed on the data of these matrices to find out if those subjects who had shown positive learning at the two-alternative forced-choice tasks as regards the voicing contrast, would also achieve positive results on the same contrast in a four-alternative forced-voice task with four competing stop consonants. These four consonants differed both in the voicing feature and in the place of articulation feature.

After the learning results had been analyzed, the data from the same matrices were used to examine whether the response patterns of the six auditorily handicapped experimental subjects pointed to independent transmission of the voicing features and place features. We already pointed out in chapter 2.2. that it is an advantage when during auditory speech training the distinctive features are perceived independent of acoustic context.

5.2.3.1. The learning results with the /t-d-p-b/ lists.

For the analysis of the learning results, based on the pooled responses of the /t-d-p-b/ tests of six experimental subjects, the phoneme matrices of table 16 were converted into separate feature matrices for the voice features [+voice] and [-voice] and the place features [+alveolar] and [-alveolar] ; here, [-alveolar] meant 'bilabial'. These stimulus-response confusion matrices were expressed in relative frequencies, and there was such a matrix of the pre-training and the post-training results.

Relative frequencies were calculated about all observed hit responses and false-alarm responses, respectively. To determine the relative frequencies of the voice responses, each correct acceptance of a voicing feature was considered a hit, irrespective of the place of articulation of the relevant decision. E.g., both /t/ and /p/ responses were considered hits in the case of stimulus /t/; and both /b/ and /d/ were considered hits with stimulus /b/. In the same way the incorrect acceptance of a voicing feature was considered a false-alarm, irrespective of the place of articulation of the relevant phoneme. E.g., /d/ and /b/ responses were considered false-alarms in the case of stimulus /p/ or /t/; /p/ and /t/ responses were false-alarms with stimulus /b/ or /d/. When the place of articulation matrix was composed, the same procedure was used, while in this case hits and false-alarms were determined by the place of articulation features, irrespective of the voicing features of the relevant decisions. The results of the reconstruction of table 16 are given in table 17.

- Before -

VOICE

stimulus	response +	response -
+	.826	.174
-	.415	.585

PLACE

stimulus	response +	response -
+	.806	.194
-	.146	.854

- After -

VOICE

stimulus	response +	response -
+	.892	.108
-	.187	.813

PLACE

stimulus	response +	response -
+	.883	.117
-	.109	.891

Table 17: Stimulus-response matrices of the observed [+voice] and [-voice] features, and the [+place] and [-place] features, before- and after training. Hit rates and false-alarm rates are given in relative frequencies.

The relative frequencies in table 17 reflect an improved perception of all relevant voice and place features, since for each feature the hit rate has increased after training, while the false-alarm rate has decreased simultaneously.
In order to determine the positive changes in the recognition of the relevant features by means of the observed hit rates and false-alarm rates, the measure of signal detectability was computed for both feature contrasts in the form of the Pollack-Norman index of performance, A' (cf. section 4.3.3.4). After training the index of performance changed from .80 to .91 for the voicing features, and from .90 to .94 for the place features. The positive changes in recognizing the place of articulation features without a previous specific training is perhaps based on the fact that during the presentation of the pre-training and post-training lists the listeners were frequently exposed to different voicing as well as to different place of articulation features (cf. section 4.2. about the exposure phenomenon for a possible explanation of a spontaneous shift in the perception of small critical differences in the speech signal).
There is a relatively considerable positive change of the index of performance for the voicing contrast. This supports the hypothesis that after specific auditory training the perception of the [+voice] and [-voice] features shows an improvement not only with a two-alternative task but also with a more extensive and varied task.

5.2.3.2. Analysis of feature independence.

In section 5.2.2.1. it had been demonstrated by means of the two-alternative task that the auditorily handicapped experimental subjects had learnt to perceive the voicing features better after auditory training, irrespective of the place of articulation of the phoneme they were part of. This conclusion strongly suggested that the voicing features could be perceived and trained independent of the place features.
To find out whether this was also true for the four-alternative task where voice features and place features competed at each decision, the observed data were used from the matrix of table 16. These data had been converted to relative frequencies per place feature and voice feature in table 17.
For each cell of the pre- and post-training confusion matrices, the predicted absolute response frequency was calculated by means of these relative frequencies in such a way that this value corresponded with an independent perception of the voice features and place features. E.g., to calculate the predicted frequency of *f(stimulus /t/ ⟶ response /p/)* we started from the formula: *f(stimulus /t/ ⟶ response /p/)* = *P(stimulus /-voice/ ⟶ response /-voice/)* x *P(stimulus /+place/ ⟶ response /-place/)* x *384*.
For immediate comparison, the results of the predicted matrices are given in table 18 together with the results of the observed matrices.

		response				
		/t/	/d/	/p/	/b/	N
stimulus	/t/	181 186	128 127	44 30	31 41	384
	/d/	54 49	256 257	13 27	62 51	384
	/p/	33 24	23 28	192 209	136 123	384
	/b/	10 21	46 39	57 37	271 287	384

BEFORE

		response				
		/t/	/d/	/p/	/b/	N
stimulus	/t/	276 280	63 47	37 34	8 23	384
	/d/	37 31	302 320	5 10	40 23	384
	/p/	34 28	8 10	278 282	64 64	384
	/b/	5 15	37 31	37 27	305 311	384

AFTER

Table 18: Predicted (upper cell) and observed (lower cell) stimulus-response patterns of six experimental subjects before and after auditory training.

The next matter to be looked into was whether the observed and the predicted matrices differed significantly. For this problem a X^2-goodness-of-fit-test was applied on the assumption that, in the case of $X^2 = 0$, there was an absolute independence between place features and voice features. There appeared to be a significant difference between the predicted and the observed matrices both before and after training (viz. $X^2 = 52{,}74$; df=9; two-tailed $p < .001$ and $X^2 = 72{,}17$; df=9; two-tailed $p < .001$, respectively). Therefore, statistically there was no complete correspondence between the predicted and the observed cell values.
Further inspection of the values per single cell proved that in 13 out of 16 cells of the pre-training matrix there was no highly significant difference between predicted and observed frequencies. Three cells showed highly significant differences, viz. (stimulus /d/ - response /p/), (stimulus /b/ - response /t/), and (stimulus /b/ - response /p/).
In the post-training matrix there was no highly significant difference between predicted and observed frequencies in 13 out of the 16 cells either. There were significant differences in three cells, viz. (stimulus /t/ - response /b/), (stimulus /d/ - response /b/), and (stimulus /b/ - response /t/).
All cells with significant differences between predicted and observed frequencies concerned false-alarm responses with a relatively low frequency. Therefore it appeared that, both in the pre-training and in the post-training matrix, differences within relatively small response categories of only three cells highly contributed to producing a significant X^2. In the pre-training matrix these cells contributed 34,20 to $X^2 = 52{,}74$; in the post-training matrix they contributed 55,35 to $X^2 = 72{,}17$.*)

The large contribution of a relatively small number of cells with a low cell frequency to the total X^2-values, and the entire lack of any system in these

---

*) *For more arguments against the statistic validity of the two* X^2*-values, see also appendix D.*

false-alarms make it clear that the established deviations are very limited and theoretically insignificant. A conspicuous fact, however, is the very large correspondence between predicted and observed data. Therefore, our conclusion is still that the perception of voice and place features shows a high degree of mutual independence.

CHAPTER 6

CONCLUSIONS AND DISCUSSION

This study was intended to improve a small aspect of the specific human domain of speech, viz. the perception of the distinctive features of voicing. Within a formal experimental learning situation, a number of hypotheses were confirmed concerning the effect of auditory training with auditorily handicapped children on the perception of the voicing features in stop consonants.

The general conclusion based on the results of this study is that, by means of a highly specific form of auditory training, three out of four hearing-impaired children as well as three dysphatic children who acted as experimental subjects showed improvement in learning to identify the voicing distinction in initial stop consonants. It was observed that this improvement was significantly greater in the group of auditorily handicapped experimental subjects than in a group of hearing-impaired control subjects who had not received this specific training. This fact justifies the conclusion that this effect was realized under the influence of specific auditory training.

In the first place, it appeared that a perceptual contrast which was manipulated artificially by means of an acoustic variable in the speech signal, the Voice Onset Time, could be categorized better after this training in accordance with the basic assumptions of categorical speech perception. VOT is the most important cue in the speech signal to evoke the adequate perception of a voiced or voiceless initial stop consonant. However, in this stage of our research it was not yet certain whether the improved ability of the auditorily handicapped subjects to categorize on the basis of the variable VOT-cue concerned an acoustic or a phonetic mode of processing.In other words, it could not yet be concluded whether merely acoustic differences were categorically perceived better after training or differences between the phonetic voicing characteristics, viz. [+voice] and [-voice] .

Furthermore, it was not yet certain whether the observed improvement also had a wider linguistic effect, and whether the perceptual information of the voicing distinction acquired by training would also play an important part in processing other stop consonants, even when occurring in varied vowel contexts.

An indication that the process of learning had rather had a phonetic (and consequently, an abstract-linguistic) effect and not merely a limited contextual acoustic effect, was given by the results of six out of seven experimental subjects, on account of two-alternative forced-choice tasks with the voicing contrast as a critical variable. These results suggested that there was actual carry-over of perceptual information, obtained via a limited pair of stimuli, to acoustically more varied stimuli in various vowel environments. In most cases, improved perception of the contrasting voicing features that were part of two given homorganic initial pre-vocal stop consonants also appeared to result in an improved perception of these features, if they belonged to the same phonemes in varied vowel contexts. It also resulted in improved perception of the voicing distinction if stop cognates had the same position in the word as the stimuli used during training but had not been produced homorganically, in other words if they differed both in place of articulation and in vowel context from the trained stimuli. Therefore, the voicing features did not seem to be context bound as regards to the vowel context and the place of articulation of the phonemes.

Training based on a limited linguistic unit had an obvious and relatively wide effect (cf. also McReynolds & Engmans, 1975; Bennet & Ling, 1977); at the same time these results suggest the existence of context independence of the learned features so that the economic gain we expected of such perceptual training seemed to be realized. In this way the linguistic relevance of training by means of limited subphonemic units with auditorily handicapped children seemed to have been demonstrated, at least as regards the voicing features.

There were more indications that it was possible to train the perception of voicing features independent of their contexts; these indications were based on the results obtained by means of the [± voice] x [± place] four-alternative forced-choice task with the /b-p-d-t/ lists. The six experimental subjects who had achieved positive performances with the two-choice tasks also clearly appeared to show systematic improvement in the perception of the voicing features with the four-choice tasks. In spite of some unaccountable noise in their response patterns, their performances showed sufficiently that voice features and place features were perceived independently.

The mainly negative post-training performances of experimental subject 3 (cf. table 14) point to a perceptual processing during the multiple choice tasks which is the opposite of the results of the other six experimental subjects, cf. supra. This very seriously hearing-impaired girl reached the criterion of training with two VOT-stimuli relatively fast, and subsequently she could identify the relevant series of VOT-stimuli adequately in a categorical mode. But she did not succeed in identifying the lists spoken with natural voice distinctions in varied contexts better after training than before. In spite of training with two alveolar VOT-stimuli she did not show any positive linguistic learning in the form of sufficient carry-over of acquired voicing information to homorganically produced stop consonants in varied vowel environments. She did not even adequately identify bilabial stops that corresponded as regards place of articulation with the a priori correctly categorized bilabial VOT-series.

Possibly this experimental subject only reacted to relevant acoustic temporal differences during the discrimination training with two extreme voice stimuli, and also during the identification task with the two VOT-series. Or she may only have reacted to a subjectively experienced presence or absence of specific acoustic information without converting the acoustic signal into a phonetic message. In the latter case it might be called pseudo-categorical speech processing, which as a mode of perception did not yield results when the same phonetic information had to be extracted from more varied acoustic signals (for categorical perception of non-speech stimuli after training or based on experience, cf. also Burns & Ward, 1974; Lane, 1965; Locke & Kellar, 1973).

With a psycho-physical variant of Ades'theory (1977), the difference in results between the various tasks of subject 3 could be connected with the difference of contents between the respective tasks and the given stimuli. The first task, i.e. the training with two VOT-stimuli and subsequently the identification of the post-training VOT-series, was based on the very restricted context of two-alternative forced-choice trials with stimuli that always differed only in one dimension, viz. in VOT. In terms of memory noise, this context could be considered relatively noise-free. The result could be that experimental subject 3 did not show any memory deficit and performed well on this task.

The second task, i.e. identifying consonants in two-alternative forced-choice trials which differed from the training stimuli as regards their vowel environment alone or the vowel context and the place of articulation together, was based on a less restricted context because of the existence of more perceptual dimensions within the stimuli. This context also contained more memory noise as a result of the presence of more perceptual dimensions in the stimuli. A wider context of the task might cause a memory deficit in this subject on account of a larger rate of memory noise. During auditory training and speech therapy with seriously hearing-impaired children, memory problems often manifest themselves with regard to complex auditory speech stimuli; yet this explanation remains speculative, and we do not presume to declare that subject 3 would not be able to learn to process the time-coded voicing cue in a speech mode. From her speech performance (cf. table 5) it also appeared that, in spite of great problems with a correct articulation, she could rely on sufficient auditory feed-back to generate time-coded phonemes.

One subject, no. 4, was able to generalize voicing information in spite of an even more serious hearing loss than was the case with subject 3. This fact is not an exception in the domain of educational audiology, since in many cases the audiometric threshold does not appear to be a reliable criterion to predict at what level hearing-impaired children will perform with their residual hearing (Ling, 1978).

It proves to be an important result of this study that both children with a disturbance of auditory signal transmission and children with a disorder of auditory signal procession appeared to benefit by specific auditory training.

Besides the therapeutic consequences of the choice of VOT-stimuli as regards auditory training of hearing-impaired children, the application of a series of VOT-stimuli with dysphatic children gave sufficient indication that these stimuli may also have a neuro-psychological diagnostic value. E.g. audiometrically normal-hearing children with a developmental dysphasia appeared unable to perceive the time coded VOT-stimuli adequately. This perceptual disability in these children suggests a possible presence of a disturbance in the higher-order, central auditory processing of specific critical time relations in the speech signal (cf. also Lowe & Campbell, 1965; Blumstein et al., 1977a; Basso et al., 1977; Tallal & Piercy, 1978).

This research produced many positive results. Still, countless questions remained unanswered. These problems did not get a chance to be solved because the hearing-impaired young subjects could not be exposed to lengthy monotonous listening experiments endlessly. Some of these unsolved problems are mentioned here.

The experiment took place in formal laboratory surroundings. During the listening tests a maximal claim was laid to the attention, listening readiness (Whetnall & Fry, 1964; Griffiths, 1967), and perceptual readiness (Bruner, 1957) of each single experimental subject as regards the VOT-cue. It remains to be seen if this commanded listening behavior would also manifest itself spontaneously after the experiment in the natural listening situation with the carelessly cued, blurring speech of normal-hearing communication partners (Lisker & Abramson, 1967). Would not this acquired listening behavior disappear again in the everyday situation for lack of sufficient relevant auditory information and develop into permanent extinction of hearing reactions to acoustic cues that are perhaps perceived clearly in an ideal learning situation only? (cf. Mark & Hardy, 1958, who claimed such extinction in young hearing-impaired children; cf. also van Uden, 1974, about the phenomenon of 'deafening' in seriously hearing-impaired children). This question is not academic if we also take into account that during the listening tasks with the hearing-impaired experimental subjects an optimal linear sound amplification was used, whereas they wore their more limited individual hearing-aids after the experiment.

Another restriction of our research concerns the context and variations within which the VOT-cue was trained. They had been strongly reduced on purpose (cf. Liberman in Stark, 1974; Bennet & Ling, 1977). During the perceptual learning research, e.g., no semantic or syntactic context-redundancy, which is often relevant, was used. No perceptual carry-over of the trained voicing principles to stop consonants in connected speech was examined. Neither was carry-over to non-initial stop consonants examined. With hearing-impaired subjects, the question is if similar learning effects as have been achieved with the relatively context-stable VOT-cue (Lisker & Abramson, 1967) can also be expected if an auditory training is given with more highly coded stop consonants, which for example differ in place of articulation and are often perceived more deficiently than the voicing distinction on account of their high degree of complicated encodedness and acoustic contextual variability (Blumstein et al., 1977b).

Will all other relevant distinctive features that form a contrast act like the voicing features in this research if they are trained in a similar way? And is there transfer of learned voicing information from stop consonants to phonemes that are realized with a different manner of production, e.g. to voiced and voiceless fricatives? The latter problem is not hypothetical since Miller (1977) recently pointed out that the voicing phoneme boundary varies not only in function of place of articulation but also in function of manner of production.

Another question that remained unsolved in this experiment concerns the problem of the listener's adjustment to the voice characteristics of the speaker. For it is a well-known fact in educational audiology that many hearing-impaired children have difficulty in adapting themselves auditorily to the characteristics of voice and speech of a strange speaker, whereas they have fewer problems in their contacts with a speaker they have known for some time. During the experiment all speech stimuli of dependent and independent variables were produced by the same male speaker. It might be assumed that the hearing-impaired subjects gradually adapted themselves easier during frequent listening to the acoustic characteristics of this speaker's voice than if they had been confronted with a variety of different voices. The question is, therefore, if the observed transfer effects on the basis of speech stimuli of one speaker will also be observed if these stimuli are produced by different speakers.

Besides the restrictions concerning the design of the experiment some more can be adduced.

No research has been done with the experimental subjects about the long-term retention of perceptual VOT-information. Although the post-training data were collected after at least one day after the termination of the training, which according to the rules can be considered a long-term re-examination, it would be interesting to examine after an even much longer time how high the rate of retention would be then. From the results of the exposure phenomenon experiment it appeared that a normal-hearing subject can still perceive after a very long time, without further exposure, a perceptual cue which he has once learnt to detect in the speech signal. But it is known from speech-therapeutic and educational-audiological experience with seriously hearing-impaired children that without regular auditory and articulatory training, there will soon be a decrease of achievements in auditory perception and in articulation.

At last, in the case of the two very seriously hearing-impaired subjects it is also possible that they did not only react to the time-coded acoustic information of the VOT-signal. In consequence of the necessary high sound intensity of the presented stimuli they might have reacted to cutaneous time information that arose from the vibration of the headphones. Via the vibro-tactile modality seriously hearing-impaired children are able to identify the different underlying acoustic patterns of the voicing features of a consonant (Kloster Jensen & Jussen, 1970; Schulte, 1979).

In spite of all these restrictions of the experiment it can be claimed that specific training with VOT-stimuli has had a fundamental, extensive effect on the perception of the voicing distinction in hearing-impaired children. Yet the ultimate conclusion of this research is not that owing to the positive results (e.g. owing to the use of modern technological devices), auditory training of speech will have to be directed more intensively in future to the use of synthetic or technically manipulated speech stimuli. In certain circumstances such stimuli make it possible to get a better insight into specific fundamental perceptual processes or into their modifications, as has appeared from this research.

On account of the results of this research some suggestions can be made about influencing the auditory abilities of young children with a hearing loss.

In the first place we are of opinion that, as a result of the extreme specificity of the speech signal and the speech processing, young hearing-impaired children should be confronted with speech as early and as much as possible, and be enabled to perceive the various complicated relevant speech cues to the best of their ability. The results of the experiments in speech perception with very young children by Eimas et al. (1971) and many others (cf. section 2.3) provide strong arguments for the earliest possible post-natal intervention in the case of children that are congenitally hearing-impaired or have acquired a hearing loss at a very early age. Early detection of the hearing disorder and timely adjustment of adequate hearing-aids in the pre-verbal period, and at least before the end of the first year of their lives, are considered a first requisite and the beginning of every auditory approach, since it is evident that at this stage aural speech abilities develop mainly on the basis of exposure to spoken language (Griffiths, 1967; Pollack, 1970; Whetnall & Fry, 1971; Beebe, 1978; Ling, 1978; Löwe, 1977). During such exposure infants are not passive recipients of outside stimuli. In a number of revealing experiments with children in their pre-verbal stage, in which automatic controlled audio-reinforcement was used in a natural environment, Friedlander showed (1968) e.g. that infants listen very selectively even in the first year of their lives and can recognize several dimensions from the range of human voice variables. These children show a clear listening preference for various configurations of normal voices; and their attention is raised for speech with high information and low redundancy (Friedlander, 1970).

A recent study in a technologically, medically and pedagogically highly developed country like Germany revealed that timely intervention with young hearing-impaired children is still a desire rather than a fact, and that little improvement can be seen in this field, in spite many years of scientific progress and instruction (Mr. Hartman, 1977; Mrs. Hartman, 1977; Löwe, 1977).

Never can it be emphasized enough that hearing results of children who have learnt to use their residual hearing in their early youth by timely help are mostly superior to those of comparable children who did not have this opportunity (Ling, 1978).

Apart from the positive effect on the hearing function, early detection of the hearing disorder and timely adjustment of hearing-aids appear also to be closely related to the speech-readiness, the verbal communication pattern and the degree to which a child's speech is intelligible to strangers at a later age (Jensema & Trybus, 1978).

In many cases it appears that post-natal auditory experience by means of an adequate hearing-aid and in the natural situation of life is sufficient to learn to recognize the most relevant speech-cues as well as possible, even without formal training. The reason is that the central nervous system of a hearing-impaired child is tuned by nature to process spoken language patterns (Fry, 1966, 1975; Ling & Ling, 1978).

We suggest, therefore, that in the very earliest stages of speech acquisition the hearing-impaired children, just like the normal-hearing ones, should be enabled to learn to process speech to the best of their ability rather by overhearing or subconscious listening than by formal parent instruction (Murphy, 1979; cf. also the result of the exposure experiment in section 4.2). Moreover, we agree with Rees (1973) that, in principle, a linguistic problem can best be coped with in a natural way and with specifically linguistic material.

In the ideal circumstances of a timely, adequate intervention, this kind of approach shows resemblance to Gibson's views on perceptual learning (1969). This view is that perceptual learning in lifelike situations is typically self-regulated and can hardly be influenced by external reinforcement.

In this early stage reinforcement does influence the so-called perceptual state, or - in pedo-audiological terms - the listening readiness (Ewing & Ewing, 1958). In this way early extinction of hearing reactions or a state of complete or partial permanent attention deficit can be prevented.

The first part of our suggestions in this section refer mainly to early otologic and educational-audiological intervention. Another primary task of educators and parents together is to develop the listening behavior of the young hearing-impaired child, until an optimum condition of hearing-alertness is realized, on which further profitable auditory training can be based.

As far as we can see, there is only direct profit of applying limited speech-units in a formal, unisensory auditory training and with hard-wire speech amplification systems in a later stage of the hearing-impaired child. In that stage a child is more or less able to cope with a structured learning situation and to concentrate his focus of attention in an operant learning situation to a cue which is not directly relevant to him but which is critical as regards a specific speech signal. This kind of training should be considered only a necessary completion of auditory experience and skills which the child has acquired in a real life situation. This training aims at gathering maximal information from minimal auditory cues.

This way of specific perceptual learning does not essentially differ from Gibson's description of the principles of a perceptual learning process (Gibson, 1969; pag. 77):

> *"Perceptual learning is an increase in the ability of an organism to get information from its environment, as a result of practice with the array of stimulation provided by the environment. This definition implies that there are potential variables of stimuli which are not differentiated within the mass of impinging stimulation, but which may be, given the proper conditions of exposure and practice. As they are differentiated, the resulting perceptions become more specific with respect to stimulation, that is, in greater correspondence with it. There is a change in what the organism can respond to. The change is not acquisition or substitution of a new response to stimulation previously responded to in some other way, but is rather responding in any discriminating way to a variable of stimulation not responded to previously. The criterion of perceptual learning is thus an increase in specificity. What is learned can be described as detection of properties, patterns, and distinctive features."*

It has appeared during this research that perceptual learning to distinguish and to identify certain speech sounds by auditorily handicapped children can be explained on account of processing relevant, relatively independent linguistic features that are based on critical acoustic patterns in the speech signal.
Just as in many previous experiments with normal-hearing subjects, it has not become clear in this research how the trained children have overcome the gap between non-invariant speech-cues and the perceptual constancy of voiced and voiceless phonemes. Neither was it the purpose of this research to find an explanation. In spite of their hearing-disorder these children proved to be able to more adequate and stabler feature processing by means of a training with limited stimuli; it is this fact we consider linguistically relevant. And as long as there is not any definite model of speech perception, it can be claimed within the context of this experiment that the experimental subjects while being trained 'went beyond the information given' (Bruner, 1973).

SUMMARY

In this research the effect of auditory training with hearing-impaired and dysphatic children was studied. This training was done by means of a critical acoustic property that caused an initial stop consonant to be perceived either as [voiced] or as [voiceless]. This variable property is called Voice Onset Time (VOT). VOT is defined as the interval between the burst explosion caused by the sudden opening of a specific part of the vocal tract resulting in a release of accumulated air on the one hand, and the beginning of vocal cord vibration on the other hand.

First, we examined whether auditorily handicapped children who could not discriminate between voiced and voiceless consonants could learn to do so by means of special training. This training was restricted to the voicing contrast between two specific plosives occurring before the same vowel. Next we examined whether auditorily handicapped children who had learnt to distinguish the voicing contrast between these two consonants could also perceive the same contrast better when the plosives occurred before different vowels or were produced in another place of articulation of the speech mechanism.

The latter question concerned the problem whether the constancy in perceiving phonemes or phoneme features can be learned in a relatively simple way in spite of the frequently varied and context dependend acoustic patterns in which they are produced. This problem was essential for the question whether auditory training with specific speech stimuli has a generalizing effect on a hearing-impaired child's perception. This question was also relevant in connection with certain opinions about the specific perception of phonemes in speech. If learning to distinguish a perceptual difference between the distinctive features [voiced] and [voiceless] would be restricted to the two consonants with which training was performed, this would be of minimal importance; it would not yield general results in the speech perception of the trained children. This would imply that the two trained features were not perceived better if the consonants they were part of occurred in another vowel context or were produced in another place of articulation. In that case, only the acoustic difference between two invariable sound patterns that were perceived as one whole would have been learned. Consequently, phoneme constancy in speech perception would have been excluded.

But it would be more efficient if learning to discriminate the distinctive features of [voiced] and [voiceless] in two specific consonants was not restricted to these two trained consonants. In that case, the difference in voicing would also be

perceived if the distinctive features belonged to consonants that occurred in varied vowel contexts or were produced in another place of articulation of the vocal tract. In that case a carry-over of information would have occurred to more varied contexts. In this research speech perception which enables such carry-over was considered linguistically relevant.

In chapter 1 we dealt with the problem that for the perceptual constancy of specific phonemes the same phonetic information must always be extracted from complicated and varied acoustic patterns. This will cause problems for the limited capacity of the human auditory system. An efficient decoding mechanism for speech is necessary to remain within the boundaries of that capacity. Since it is hard for many hearing-impaired subjects to decode speech adequately, the question is whether they perceive speech in a linguistically relevant way in the form of phonemes or phoneme features or as non-linguistic (e.g. holistic) sound patterns. Perceiving holistic instead of coded speech patterns is not efficient, and it demands of the listener an unlimited capacity of memory which he doesn't really have.
For the acquisition of verbal communication by auditorily handicapped children the change or the improvement of an inadequate mode of speech perception is of primary importance. There are several methods of training to develop auditory abilities in seriously hearing-impaired children. But in view of recent discoveries about the specific mode of speech perception, the effect of several of these methods is questionable. E.g., it is doubtful whether frequently applied auditory training by means of acoustic stimuli that do not originate from speech has a favourable influence on the perception of phonemes.
In our research the VOT was used as a relevant acoustic variable that was part of a complex speech signal. It was our intention to achieve a fundamental change in the perception of voiced and voiceless phonemes with auditorily handicapped children and to demonstrate this experimentally.

In chapter 2 the concept of VOT and some properties of the perceptual VOT-dimension were further defined. The voicing features of initial voiced and voiceless plosives are based on VOT-variations to a considerable degree and are perceived in a categorical mode. In this research the categorical speech perception paradigm was used; for that reason some important assumptions were explained in connection with the categorical perception of VOT-stimuli. By varying the VOT in equal steps within a consonant-vowel segment of the acoustic speech signal, a psycho-physical stimulus continuum can be constructed. The stimuli of this continuum can be identified in an identification test as partly voiced and partly voiceless. It appears that the transitional area on the VOT-continuum between the two perceptual categories is generally very small in the Dutch language. On account of this, the use of VOT-stimuli enabled us to determine exactly the perceptual phoneme boundary between the voiceless and voiced category on the VOT-continuum. In spite of the use of equal interstimulus VOT-intervals, comparisons between intra-category stimuli are perceived as the same phonemes (expressed in terms of [voiced] and [voiceless]), and inter-category comparisons are perceived as different phonemes, during a discrimination test with stimuli from the VOT-continuum. Moreover, the results of the respective comparisons between two VOT-stimuli can be reliable predicted from the results of the identifications.

Chapter 2 also dealt with the problem that for the acquisition of constancies in the perception of phonemes for each phoneme the presence of specific invariant acoustic information in the speech signal could be expected.But in fact,especially plosives in various vowel contexts are characterized by acoustic variability as a result of co-articulation effects. As invariant acoustic information is lacking in the speech signal of certain consonants it may be asked whether perceptual information about these consonants is carried over from one consonant-vowel context to another during auditory training.

Of some present speech perception models one was preferred in this research in which it is assumed that the phonemes from the coded speech signal are recovered by means of a purely auditory decoder without mediation of processes that also play a part in speech production. Such mediation is suggested in the well-known "Motor Theory of Speech Perception", but it did not give a plausible starting-point for the results of auditory training of auditorily handicapped children, who naturally have speech defects.

The problem of phoneme perception would have been solved more or less if the listener could convert acoustic information into linguistic phoneme information, e.g. in the form of distinctive features, by means of so-called 'property and feature detectors'. If these features unbound with regard to their acoustic contexts and mutually independent could be extracted from the speech signal, it would be a plausible explanation for the phenomenon of generalization in the auditory training of distinctive phoneme features. According to a number of studies, the distinctive features appear to be a psychological reality in speech perception. Some of these studies have shown that distinctive features are also relevant in the speech perception of hearing-impaired persons and that the auditory difference between voiced and voiceless consonants plays an important part in the perceptual strategy of a seriously hearing-impaired person.

At last chapter 2 dealt with the question whether the categorical perception of a relevant speech pattern could be modified. This question appeared to be important for a hearing-impaired person who on account of his defect has deviant perceptual phoneme boundaries or a decreased ability to categorize a phoneme contrast. Besides a theory which claims that speech perception is congenital and genetically pre-determined, a more plausible theory was described that starts from the principle that speech perception is to a considerable degree based on experience originating from the linguistic environment. According to this latter theory speech perception can always be modified in spite of the existence of a critical period of development at an early age. This theory maintains, therefore, that there remains a possibility of auditory training even at a later age. This was also postulated for the succes of the learning experiment in our research. At the end of chapter 2 some suggestions were made about the way a change in categorical speech perception may occur and about the way this might manifest itself during the research.

In chapter 3 the computer controlled construction of an alveolar /dɑk-tɑk/ stimulus series and a bilabial /bɑk-pɑk/ stimulus series was described.
The VOT was varied systematically in the initial part of the acoustic signal of the normally spoken syllables [bɑk] and [dɑk] by means of a segmentation procedure. In this way it was tried to produce a perceptual shift from the initial voiced consonant to an initial voiceless consonant. Next, the phoneme boundary values on the VOT-continuum were determined for both VOT-series by means of an identification test. The identifications were performed by a group of normal-hearing listeners, whose ages ranged from 4 to 40. The bilabial and alveolar stimulus series both showed the desired perceptual transition from the voiced category to the voiceless category. They also appeared to be perceived in accordance with the assumptions of the categorical speech perception. There appeared to be no systematic differences of age or sex between the phoneme boundary values, so that these values could be used later as reference data in the training research with auditorily handicapped children.
A slight but significant difference was found between the perceptual phoneme boundaries of the alveolar and the bilabial series of VOT-stimuli. This so-called place of articulation effect suggested that the VOT-perception had not been entirely independent of the perception of the place of articulation.

Chapter 4 dealt with the selection of a group of hearing-impaired children and a group of dysphasic ones of about 10. All these children had difficulty in categorically perceiving the perceptual voice contrast in one or both VOT-series. They took part in the research as experimental and control subjects.
The choice of the method that was applied during the research was based on the result of a preceding experiment that was called the "Exposure Phenomenon". For it appeared that a normal-hearing listener could not at first distinguish a contrast based on VOT-differences; later, however, he was able to distinguish this contrast by prolonged listening to the two extreme stimuli of the relevant VOT-continuum.
Each of the auditorily handicapped experimental subjects received a discrimination training with two different VOT-stimuli. These stimuli had been selected for each individual experimental subject from the VOT-series which before had been identified with the worst results. For each subject the choice of the two training stimuli was specifically directed in accordance with the deviation of his identifications. The idea was to produce a shift of the relevant phoneme boundary on the VOT-continuum or an increase of the number of correct identifications of one or both categories of

a VOT-serie. The training was based on the principles of instrumental operant conditioning. Relevant auditory feed-back was used, and practice was continued until a specific criterion had been reached or a pre-arranged limit of the number of practice sessions was exceeded. On account of a non-parametric calculation of the measure of signal detection it could be demonstrated that all subjects learned to distinguish the contrast between the two training stimuli within the arranged criterion in a satisfactory way.

In chapter 5 the effect of auditory training was described. First it was studied with each subject whether training with two specific VOT-stimuli had a positive effect on the perception of the whole alveolar or bilabial VOT-series from which these stimuli had been selected. It sometimes appeared that the identifications of the other VOT-series from which no training stimuli had been selected also deviated before the training. In that case it was examined whether the perception of this non-trained series was also influenced positively by the discrimination training with the two stimuli that differed from this series in place of articulation.
It appeared that with all subjects a positive effect of the training could be demonstrated in the perception of trained and non-trained VOT-series. The perceptual changes manifested themselves after the training in the perceptual functions of both series of stimuli, either in a quantitative improvement of categorization, or in a desired phoneme boundary shift on the VOT-continuum. These first positive results suggested that the distinctive features of [voiced] and [voiceless] in initial consonants could be perceived and trained independent of their place of articulation.
Next it was examined whether the positive effect of the training was linguistically relevant. This was done by comparing the pre-training and post-training results of three different listening tasks.
The first task consisted of a series of word pairs that differed only in the voicing contrast of the initial consonants. During this task the subjects had to point out which of the two words they had heard. The initial consonants of these so-called minimal pairs were the phonemes /d/ and /t/. There was a wide variety of vowels among the various minimal pairs. The auditory stimuli were pronounced normally and naturally.
The second task was similar to the first, but the initial consonants were now the bilabials /b/ and /p/.
During the third task the subjects had to identify an auditorily presented word by choosing from four words which were completely identical except for the different initial consonants /t-d-b-p/. The four initial phonemes within each item were characterized by the features [voiced] or [voiceless] and by the place of articulation, viz. [bilabial] or [alveolar]. The vowel contexts of the initial consonants were varied per item.
The identification results of both two-alternative tasks were first combined per voicing feature. The group of experimental subjects appeared to perceive the voicing

contrast better after training than before. Moreover, this group showed significantly more improvement in perceiving the voicing contrast than the control group which had not been trained and which had performed the two listening tasks before and after a certain latency period. Therefore, it was concluded that the perceptual improvement of the experimental group could be ascribed to the training with specifically selected VOT-stimuli.

Next, the results of the /d-t/ and /b-p/ lists were analyzed for each individual subject and per voicing feature according to a twofold criterion. This criterion was to show whether a change in perceiving the voicing distinction was linguistically relevant. According to this criterion a subject's post-training results should point to a significant positive or negative change in the perception of one or both distinctive voicing features occurring in a list; moreover, such a change should also imply a jump from a negative to a positive perceptual level of performance or vice versa. The limits between these positive and negative levels could be determined statistically per voicing feature. On the basis of this twofold criterion the identifications of one or both voicing features appeared to have improved for the whole group of experimental subjects in a linguistically relevant way. This was the case for the listening task of which the words had initial consonants that corresponded with the VOT-training stimuli as regards the place of articulation. The results of the group of experimental subjects also showed that there had been a linguistically relevant and positive change in the identifications of one or both voicing features in the listening task where the initial stops differed from the training stimuli in place of articulation. No experimental subject showed any relevant negative change. These results showed once more that perceptual learning of the voicing distinction in initial consonants was independent of the place of articulation and the vowel context of the respective consonants.

In the case of one experimental subject there were no systematic positive changes in the results of the two-alternative tasks. These results were excluded from further analysis of the /t-d-p-b/ four-alternative identification task.

The combined results of the /t-d-p-b/ tasks of the other six experimental subjects, who had shown positive and linguistically relevant changes in their identifications in the other two tasks, were subjected to a distinctive feature analysis. Both the distinctive voicing features of [voiced] and [voiceless] and the distinctive place features of [alveolar] and [bilabial] were involved in this analysis. A non-parametric calculation of the measure of signal detection was then performed based on the correct as well as the incorrect acceptances of the distinctive features; the index of performance showed that after training there had been a relatively considerable improvement in the perception of the voicing contrast, and a relatively small improvement in the perception of the place of articulation contrast. Consequently we concluded that after specific auditory training there was an improvement, not only in the results of a two-alternative task, but also in those of a more extensive and more varied task.

The relatively small improvement in the perception of the place of articulation

contrast, which had not been involved in the specific training, could be interpreted more or less speculatively on the ground of the theory of the 'Exposure Phenomenon'. At last, by means of the responses of the /t-d-p-b/ four-alternative task it was examined whether the two voicing features and the two place features were perceived mutually independent, before and after training. To this end the observed stimulus-response confusion matrix was compared with the predicted confusion matrix. The hypothesis was that in the case of complete correspondence between the two matrices it could be concluded that voice and place features were perceived mutually independent. For this purpose the predicted confusion matrix was calculated by means of the observed confusion patterns of the four phonemes on the basis of an independent perception of the distinctive voice and place features.
Since a statistic test demonstrated that there were differences between the predicted and the observed confusion matrices, the correspondence was statistically incomplete. But close inspection of all separate stimulus-response categories demonstrated that the differences between the observed and the predicted confusion matrices could mainly be ascribed to response-noise of little theoretical importance within a small number of confusion categories. It also demonstrated that the correspondence of the other categories was great. This fact, as well as the positive results of the learning experiment with the VOT-stimuli and the demonstrated positive linguistic effect in the two-alternative tasks, gave sufficient rise to maintain the model of an independent processing of the distinctive voice and place features.

In chapter 6 the conclusions about the purpose and the results of our research were summarized. The consequence of the positive results of auditory training by means of the distinctive features of [voiced] and [voiceless] were explained. In accordance with the hypothesis it appeared that a generalizing effect was produced by a relatively limited training with variable acoustic stimuli which were part of a speech signal and which were critical of the perception of the distinctive voicing features. This effect was based on the fact that the sub-phonemic voicing features were processed as perceptually relevant, independent speech units. This perception was probably produced by means of an auditory decoding mechanism which could extract the features from the acoustic signal without the mediation of speech processes. The conclusions suggested that training of speech perception by means of distinctive features is efficient from an educational-audiological point of view.
Chapter 6 ended with a discussion about some limitations of our research. These limitations appeared to be inevitable on account of the listening tests and the nature of the subjects.

In dit onderzoek werd het effekt nagegaan van een auditieve training bij gehoorgestoorde- en dysfatische kinderen. Deze training werd uitgevoerd met behulp van een kritische akoestische eigenschap die de waarneming van de distinktieve kenmerken [stemhebbend] en [stemloos] in eksplosieve medeklinkers aan het begin van een lettergreep teweegbracht. Deze variabele eigenschap wordt Voice-Onset-Time (VOT) genoemd. De VOT wordt gedefinieerd als de tijdsduur tussen enerzijds de eksplosie die het gevolg is van het plotseling ontsluiten van een bepaald deel van het spraakmechanisme, tengevolge waarvan opeengehoopte lucht ontsnapt,en anderzijds de aanvang van stembandvibratie.

In de eerste plaats werd onderzocht of auditief gehandicapte kinderen, die stemhebbende- en stemloze medeklinkers niet goed van elkaar konden onderscheiden, dit via een speciale training konden leren. Deze training beperkte zich tot het stemkontrast tussen twee bepaalde eksplosieve medeklinkers die voor dezelfde klinker stonden. Vervolgens werd nagegaan of de auditief gehandicapte kinderen die het stemkontrast tussen deze twee medeklinkers beter leerden onderscheiden, hetzelfde kontrast ook beter konden waarnemen als de eksplosieve medeklinkers voor verschillende klinkers stonden of op een andere artikulatieplaats van het spraakmechanisme waren geproduceerd.

Deze tweede vraag betrof het probleem of de konstantie in de waarneming van fonemen of foneemeigenschappen op betrekkelijk eenvoudige manier kan worden aangeleerd, ondanks de vaak gevarieerde- en van de kontekst afhankelijke akoestische patronen waarmee ze worden geproduceerd. Dit probleem was fundamenteel voor de vraag of auditieve training met bepaalde spraakstimuli een generaliserend effekt heeft op de waarneming van een gehoorgestoord kind. Verder was deze vraag relevant in verband met bepaalde opvattingen over de specifieke waarneming van fonemen in de spraak.

Als het aanleren van een perceptueel verschil tussen de distinktieve kenmerken [stemhebbend] en [stemloos] beperkt zou blijven tot de twee medeklinkers waarmee de training werd gegeven, zou dit slechts van minimale betekenis zijn en geen algemeen profijt opleveren voor de spraakwaarneming van de getrainde kinderen. Dit zou betekenen dat de twee getrainde kenmerken niet beter werden waargenomen als de medeklinkers, waarvan ze deel uitmaakten, in een andere klinker-omgeving stonden of op een andere artikulatieplaats werden geproduceerd. In dat geval zou slechts het akoestische onderscheid tussen twee onveranderlijke en een als één geheel

waargenomen geluidspatroon zijn aangeleerd. Daarmee zou foneemkonstantie in de waarneming van spraak uitgesloten zijn.
Maar het zou efficiënter zijn als het leren onderscheiden van de distinktieve kenmerken [stemhebbend] en [stemloos] in twee bepaalde medeklinkers zich niet beperkte tot de twee getrainde medeklinkers. Het stemonderscheid zou dan ook worden waargenomen als de distinktieve kenmerken behoorden tot medeklinkers die in gevarieerde klinkerkonteksten stonden of op een andere artikulatieplaats van het spraakmechanisme werden geproduceerd. In dat geval zou een overdracht van informatie naar gevarieerde konteksten hebben plaatsgevonden. De waarneming van spraak die een dergelijke overdracht mogelijk maakt werd in dit onderzoek linguistisch relevant genoemd.

In hoofdstuk 1 werd het probleem behandeld dat voor de waarnemingskonstantie van bepaalde fonemen steeds dezelfde fonetische informatie uit gekompliceerde- en gevarieerde akoestische patronen moet worden gehaald. Dit zal problemen opleveren voor de beperkte kapaciteit van het auditieve systeem van de mens. Een efficiënt dekoderingsmechanisme voor spraak is vereist om die kapaciteit niet te overschrijden. Aangezien veel auditief gehandicapten moeite hebben met het adekwaat dekoderen van spraak, is het de vraag of zij spraak op linguistisch relevante wijze in de vorm van fonemen of foneemeigenschappen waarnemen, dan wel als niet-linguistische (bijvoorbeeld holistische-) geluidspatronen. De waarneming van holistische- in plaats van gekodeerde spraakpatronen is niet efficiënt en vraagt van de luisteraar een onbeperkte geheugenkapaciteit, waarover hij in feite niet beschikt. Voor het aanleren van verbale kommunikatie bij auditief gehandicapte kinderen is de verandering of verbetering van een inadekwate spraakwaarneming van primair belang. Voor de ontwikkeling van auditieve vaardigheden bij ernstig gehoorgestoorde kinderen bestaan verschillende trainingsmethoden. Maar in verband met recente bevindingen wat betreft de specifieke wijze van spraakwaarneming, kan aan het effekt van een aantal van deze methoden worden getwijfeld. Het is bijvoorbeeld dubieus of de dikwijls toegepaste auditieve training met akoestische stimuli die niet van spraak afkomstig zijn, een gunstige invloed heeft op de waarneming van fonemen.
In dit onderzoek werd de VOT gebruikt als een relevante akoestische variabele en deel uitmakend van een kompleks spraaksignaal. De bedoeling was om bij auditief gehandicapte kinderen een fundamentele verandering in de waarneming van stemhebbende- en stemloze fonemen tot stand te brengen en dit proefondervindelijk aan te tonen.

In hoofdstuk 2 werden het begrip'VOT'en enkele eigenschappen van de perceptuele VOT-dimensie nader omschreven. De stemkenmerken van stemhebbende- en stemloze eksplosieve medeklinkers aan het begin van een woord berusten in belangrijke mate op VOT-variaties en worden op kategorische wijze waargenomen. Omdat in het onderzoek gebruik werd gemaakt van het kategorisch spraakperceptie-model, werden enkele belangrijke assumpties in verband met de kategorische waarneming van VOT-stimuli uiteengezet.

Door binnen een medeklinker-klinker segment van het akoestische spraaksignaal de VOT in gelijke stappen te variëren, kan een psychofysisch stimuluskontinuum worden gekonstrueerd. Hiervan worden de stimuli bij een identifikatietest voor een deel als stemhebbend en voor een deel als stemloos geïdentificeerd. Het blijkt dat het overgangsgebied op het VOT-kontinuum tussen beide waarnemingskategorieën voor het Nederlands doorgaans zeer klein is. Op grond hiervan is het met behulp van de VOT-stimuli mogelijk om voor iedere individuele luisteraar exact de perceptuele foneemgrens tussen de stemloze- en stemhebbende kategorie op het VOT-kontinuum te bepalen. Bij een diskriminatietest met stimuli uit het VOT-kontinuum worden, ondanks de toepassing van gelijke interstimulus-intervallen, vergelijkingen tussen binnenkategorie stimuli als dezelfde fonemen waargenomen (uitgedrukt in termen van stemhebbendheid of stemloosheid) en tussen-kategorie vergelijkingen als verschillende fonemen. Bovendien kunnen de resultaten van de respektievelijke vergelijkingen tussen de VOT-stimuli betrouwbaar worden voorspeld uit de resultaten van de identifikaties.

Hoofdstuk 2 behandelde verder het probleem dat men voor het verwerven van konstanties in de waarneming van fonemen voor ieder foneem de aanwezigheid van specifieke- en onveranderlijke akoestische informatie in het spraaksignaal zou verwachten. Maar in werkelijkheid worden vooral de eksplosieve medeklinkers in verschillende klinkerkonteksten tengevolge van koartikulatie-effekten gekenmerkt door akoestische variabiliteit. Aangezien bij bepaalde medeklinkers invariante akoestische informatie in het spraaksignaal ontbreekt ten gevolge van koartikulatie-effekten, kan de vraag worden gesteld of bij een auditieve training perceptuele informatie betreffende deze medeklinkers wordt overgedragen van de ene- naar de andere medeklinker-klinker kontekst.

Uit verschillende spraakperceptiemodellen werd in dit onderzoek de voorkeur gegeven aan een model waarbij de fonemen in het gekodeerde spraaksignaal worden ontdekt met behulp van een auditief dekoderingsmechanisme zonder bemiddeling van processen die ook een rol spelen bij de spraakproduktie. Een dergelijke bemiddeling wordt in de bekende 'Motor Theory of Speech Perception' wel gesuggereerd, maar bood voor de resultaten van een auditieve training met auditief gehandicapte kinderen, die per definitie spraakafwijkingen hebben, geen plausibel uitgangspunt.

Het probleem van de foneemwaarneming zou min of meer opgelost zijn indien de luisteraar door middel van zogenaamde 'property' en 'feature' detektoren de akoestische informatie om zou kunnen zetten in linguistische foneem-informatie, bijvoorbeeld in de vorm van distinktieve kenmerken. Als deze kenmerken vrij van de akoestische kontekst en onafhankelijk van elkaar uit het spraaksignaal konden worden gehaald, zou dit een plausibele verklaring zijn voor het generaliseringsverschijnsel bij auditieve training van distinktieve foneemeigenschappen. Volgens een aantal onderzoekingen blijken de distinktieve kenmerken een psychologische realiteit in de spraakwaarneming. Uit enkele hiervan is gebleken dat ook in de spraakwaarneming van auditief gehandicapten distinktieve kenmerken perceptueel relevant zijn en dat het auditieve verschil tussen de kenmerken [stemhebbend] en [stemloos] een belangrijke rol speelt bij de

waarnemingsstrategie van een ernstig slechthorende.
Tenslotte werd in hoofdstuk 2 de vraag gesteld of de kategorische waarneming van een relevant spraakpatroon gewijzigd kon worden. Een dergelijke vraag bleek van belang voor de auditief gehandicapte die op grond van zijn gehoordefekt afwijkende perceptuele foneemgrenzen of een verminderd vermogen tot kategoriseren van een foneemkontrast vertoont. Naast een theorie die beweert dat de spraakwaarneming aangeboren en genetisch gepredetermineerd is, werd een meer aannemelijke theorie beschreven die ervan uitgaat dat de spraakperceptie in belangrijke mate op ervaring uit het linguistische milieu berust. Volgens deze laatste theorie zou, ondanks de aanwezigheid van een kritische ontwikkelingsperiode hiervoor in de vroege jeugd, de spraakwaarneming te allen tijde gewijzigd kunnen worden. Deze theorie gaat er dus vanuit dat een mogelijkheid tot auditief leren ook op latere leeftijd aanwezig blijft. Dit werd voor het slagen van de leereksperimenten in dit onderzoek ook vooropgesteld.
Hoofdstuk 2 werd besloten met enkele suggesties over de wijze waarop een verandering in de kategorische spraakwaarneming zich kan voordoen en hoe dit zich in het onderzoek zou kunnen manifesteren.

Hoofdstuk 3 beschreef de door een computer gekontroleerde konstruktie van een alveolare stimulusserie /dɔk-tɔk/ en een bilabiale stimulusserie /bɔk-pɔk/.
Door in de aanvang van het akoestische signaal van de normaal gesproken lettergrepen [bɔk] en [dɔk] de VOT met behulp van een segmentatieprocedure op systematische wijze te varieëren, werd geprobeerd om een verandering van de stemhebbend waargenomen beginmedeklinker in een stemloos waargenomen beginmedeklinker tot stand te brengen.
Vervolgens werden voor beide VOT-series de foneemgrenswaarden op het VOT-kontinuum bepaald met behulp van een identifikatietest. De identifikaties waren afkomstig van een groep normaalhorende luisteraars tussen 4 en 40 jaar oud. De bilabiale- en alveolare stimulusseries vertoonden beide de gewenste perceptuele overgang van het kenmerk [stemhebbend] naar het kenmerk [stemloos]. Zij bleken ook in overeenstemming met de assumpties van de kategorische spraakperceptie te worden waargenomen. Er bleken geen systematische leeftijd- of sekse verschillen tussen de foneemgrenswaarden te bestaan, zodat deze verder als referentiegegevens in het leeronderzoek met auditief gehandicapte kinderen konden worden gebruikt.
Een gering, maar signifikant verschil werd gevonden tussen de perceptuele foneemgrenzen van de alveolare- en de bilabiale VOT-stimulusseries. Dit zogenaamde artikulatieplaats-effekt suggereerde dat de VOT-perceptie niet geheel onafhankelijk van de perceptie van de artikulatieplaats had plaatsgevonden.

Hoofdstuk 4 behandelde de selektie van een groep slechthorende- en een groep van dysfatische kinderen van ca. 10 jaar. Deze kinderen hadden allen problemen met de kategorische waarneming van het perceptuele stemkontrast in één of beide VOT-series. Zij participeerden als eksperimentele- en kontrole proefpersonen aan het leeronderzoek.

De keuze van de methode die bij het leeronderzoek werd toegepast berustte op het resultaat van een voor-eksperiment dat het 'Exposure Phenomenon' werd genoemd. Het bleek namelijk dat één normaalhorende luisteraar een kontrast dat op VOT-verschillen berustte en dat hij aanvankelijk niet kon onderscheiden, door langdurig te luisteren naar de twee uiterste stimuli van het betreffende VOT-kontinuum, later wel kon leren onderscheiden.
De auditief gehandicapte eksperimentele proefpersonen kregen individueel een diskriminatietraining met twee verschillende VOT-stimuli. Deze stimuli waren voor iedere proefpersoon apart geselekteerd uit de VOT-serie die tevoren het slechtst geïdentificeerd bleek te zijn. De keuze van de twee trainingsstimuli was voor elke proefpersoon specifiek gericht op de aard van de afwijking van zijn identifikaties. De bedoeling hiervan was om een verschuiving van de betreffende foneemgrens op het VOT-kontinuum of een vermeerdering van het aantal korrekte identifikaties van één of beide kategorieën van een VOT-serie tot stand te brengen. De training berustte op de principes van een instrumentele-, operante konditionering. Er werd gebruik gemaakt van relevante auditieve feed-back en de training werd gekontinueerd totdat een bepaald kriterium was bereikt of een vooraf gestelde limiet van het aantal trainingsessies werd overschreden. Op grond van een niet-parametrische berekening van de mate van signaal-detektie kon worden aangetoond dat alle proefpersonen het kontrast tussen de twee leerstimuli op bevredigende wijze leerden onderscheiden binnen het gestelde kriterium.

In hoofdstuk 5 werd het effekt van de auditieve training beschreven. Eerst werd bij iedere proefpersoon onderzocht of de training met de twee specifieke VOT-stimuli een positief leereffekt had op de waarneming van de gehele alveolare- of bilabiale VOT-serie waaruit deze stimuli geselekteerd waren. Soms bleken vóór de aanvang van de training de identifikaties van de andere VOT-serie, waaruit geen trainingsstimuli waren geselekteerd, ook af te wijken. In dat geval werd onderzocht of ook de waarneming van deze niet-getrainde serie positief werd beïnvloed door de diskriminatietraining met de twee stimuli die in artikulatieplaats hiervan verschilden.
Het bleek dat er bij alle proefpersonen een positief effekt van de training in de waarneming van de getrainde- en ongetrainde VOT-series aangetoond kon worden. De perceptuele veranderingen manifesteerden zich na de training in de waarnemingsfunkties van beide stimulusseries ofwel door een kwantitatief verbeterde kategorisering ofwel door een gewenste verschuiving van de foneemgrens op het VOT-kontinuum. Deze eerste positieve resultaten suggereerden dat de distinktieve kenmerken [stemhebbend] en [stemloos] in beginkonsonanten onafhankelijk van de artikulatieplaats konden worden waargenomen en getraind.
Vervolgens werd onderzocht of het positieve effekt van de training linguistisch relevant was. Hiervoor werden de resultaten van vóór en na de training van drie verschillende luistertaken met elkaar vergeleken.
Bij de eerste taak moesten de proefpersonen van een aantal woordparen, die slechts

door het stemkontrast van de beginmedeklinkers van elkaar verschilden, aangeven welke van de twee woorden zij hadden gehoord. De beginmedeklinkers van deze zogenaamde minimale paren waren de fonemen /d/ en /t/. Er was gezorgd voor een grote klinkervariatie tussen de diverse minimale paren. Bij deze taak waren de auditieve stimuli op normale, natuurlijke wijze uitgesproken.

De tweede luistertaak hield hetzelfde in als de eerste, maar nu stonden de bilabiale medeklinkers /b/ en /p/ aan het begin van de woorden.

Bij de derde taak moesten de proefpersonen een auditief aangeboden woord identificeren door het maken van een keuze uit vier woorden die, afgezien van de verschillende beginmedeklinkers /t-d-p-b/, verder geheel identiek waren. De vier beginfonemen binnen iedere opgave werden gekenmerkt door de stemkenmerken [stemhebbend] of [stemloos] en door de artikulatieplaatsen [bilabiaal] of [alveolaar]. De klinkeromgevingen van de beginmedeklinkers waren bij de verschillende opdrachten gevarieerd.

De identifikatie-resultaten van beide twee-keuzen-taken werden eerst per stemkenmerk gekombineerd. Het bleek dat de eksperimentele groep het stemkontrast na de training beter kon waarnemen dan ervoor. Bovendien vertoonde de eksperimentele groep een signifikant grotere verbetering in het waarnemen van het stemkontrast dan de kontrole-groep die geen training had ontvangen en beide luistertaken vóór en na een bepaalde latentieperiode had uitgevoerd. Hieruit werd de konklusie getrokken dat de perceptuele verbetering van de eksperimentele groep toegeschreven kon worden aan de training met specifiek geselekteerde VOT-stimuli.

Vervolgens werden de resultaten van de /d-t/ en /b-p/ lijsten per individuele proefpersoon en per stemkenmerk geanalyseerd volgens een tweevoudig kriterium waaruit moest blijken of een verandering in het waarnemen van het perceptuele stemkontrast linguistisch relevant was. Volgens dit kriterium moesten na de training de resultaten van een proefpersoon niet alleen wijzen op een signifikante positieve- of negatieve verandering in het waarnemen van één of beide distinktieve stemkenmerken die in een lijst voorkwamen, maar moest een dergelijke verandering ook een sprong inhouden van een negatief- naar een positief prestatie-niveau van waarnemen of omgekeerd. De grenzen tussen de positieve- en negatieve prestatie-niveaux konden per stemkenmerk statistisch worden bepaald. Op grond van dit tweevoudig kriterium bleken de identifikaties van één of beide stemkenmerken voor de gehele groep van eksperimentele proefpersonen op linguistisch relevante wijze positief te zijn veranderd.

Dit gold voor de luistertaak met woorden die begonnen met een medeklinker waarvan de artikulatieplaats overeenkwam met die van de VOT-trainingsstimuli. Het bleek bovendien uit de resultaten van de groep van eksperimentele proefpersonen dat er een linguistisch relevante-, positieve verandering in de identifikaties van één of beide stemkenmerken was opgetreden voor de luistertaak waarbij de beginmedeklinkers in artikulatieplaats verschilden van de trainingsstimuli. Een relevante negatieve verandering kwam bij geen enkele eksperimentele proefpersoon voor.

Deze resultaten demonstreerden opnieuw dat de waarneming van de distinktieve kenmerken [stemhebbend] en [stemloos] in beginmedeklinkers onafhankelijk van de artikulatieplaats en vrij van de klinkerkontekst plaatsvond en kon worden getraind.

Bij één eksperimentele proefpersoon vertoonden de resultaten van de twee-keuzentaken geen systematische positieve veranderingen. Deze resultaten werden uitgesloten van verdere analyse bij de /t-d-p-b/ vier-keuzen-identifikatietaak.

De gekombineerde resultaten van de /t-d-p-b/ taak van de zes overige eksperimentele proefpersonenen, die wél positieve- en linguistisch relevante veranderingen in hun identifikaties met de andere twee taken hadden vertoond, werden aan een distinktievekenmerken-analyse onderworpen. Bij deze analyse werden zowel de distinktieve stemkenmerken [stemhebbend] en [stemloos] als de distinktieve plaatskenmerken [alveolaar] en [bilabiaal] betrokken. Bij een niet-parametrische berekening van de mate van signaal-detektie op grond van zowel de korrekte als de inkorrekte akseptaties van de foneemkenmerken, toonde de prestatieskore aan dat na de training in de waarneming van het stemkontrast een relatief grote- en in de waarneming van het artikulatieplaatskontrast een relatief kleine positieve verbetering was opgetreden. Dit leidde tot de konklusie dat na een specifieke auditieve training niet alleen een verbetering optrad in de resultaten van een twee-keuzen taak, maar eveneens in de resultaten van een meer uitgebreide- en gevarieerde taak.

De relatief kleine verbetering in de waarneming van het artikulatieplaats-kontrast, dat niet bij de specifieke training betrokken was geweest, kon min of meer spekulatief worden geïnterpreteerd op grond van de theorie van het 'Exposure Phenomenon'.

Tenslotte werd met behulp van de responsies van de /t-d-p-b/ vier-keuzen taak onderzocht of vóór en na de training de twee stemkenmerken en de twee plaatskenmerken onafhankelijk van elkaar werden waargenomen. Hiervoor werd de geobserveerde stimulus-antwoord verwarringsmatrix vergeleken met de voorspelde verwarringsmatrix. De hypothese luidde dat bij een volledige overeenkomst tussen de beide matrices gekonkludeerd kon worden dat stem- en plaatskenmerken onafhankelijk van elkaar werden waargenomen. De voorspelde verwarringsmatrix was voor dit doel met behulp van de geobserveerde verwarringspatronen van de vier fonemen berekend op basis van een onafhankelijke waarneming van de distinktieve stem- en artikulatieplaats kenmerken. Aangezien een statistische test aantoonde dat er verschillen bestonden tussen de voorspelde- en geobserveerde verwarringsmatrices bleek dus de overeenkomst statistisch niet volledig. Maar een nauwkeurige inspektie van alle afzonderlijke stimulus-antwoord kategorieën toonde aan dat de verschillen tussen de geobserveerde- en voorspelde verwarringsmatrices grotendeels konden worden toegeschreven aan responsie-ruis van geringe theoretische betekenis binnen een klein aantal verwarringskategorieën en dat de overeenkomst tussen de overige kategorieën wel groot was. Samen met de positieve resultaten van het leereksperiment met de VOT-stimuli en het aangetoonde positieve linguistische effekt bij de twee-keuzen taken, was dit aanleiding om het model van een onafhankelijke waarneming van de distinktieve stem- en artikulatieplaats kenmerken te handhaven.

In hoofdstuk 6 werden de konklusies betreffende de opzet en de resultaten van het onderzoek samengevat. De gevolgen van de positieve resultaten van een auditieve

training met behulp van de distinktieve kenmerken [stemhebbend] en [stemloos] werden uiteengezet. Overeenkomstig de gestelde hypothesen bleek een betrekkelijk beperkte training met variabele akoestische stimuli die deel uitmaakten van een spraaksignaal en kritisch waren voor de waarneming van de distinktieve stemkenmerken, een generaliserend effekt te hebben. Dit generaliserend effekt berustte op het feit dat de tot de verschillende fonemen behorende stemkenmerken als perceptueel relevante-, onafhankelijke spraakeenheden werden waargenomen. Deze waarneming kwam waarschijnlijk tot stand door middel van een auditief dekoderingsmechanisme dat de kenmerken uit het akoestische signaal kon halen zonder dat spraakprocessen hierbij een bemiddelende rol speelden. De konklusies suggereerden dat een training van de spraakwaarneming met behulp van distinktieve kenmerken vanuit leerpsychologisch oogpunt efficiënt is.
Hoofdstuk 6 werd besloten met een diskussie betreffende een aantal beperkingen van het onderzoek. Deze beperkingen bleken door de soort van luisterproeven en de aard van de proefpersonen onvermijdelijk te zijn.

REFERENCES

Abbs, J. & Sussman, H. Neurophysiological feature detectors and speech perception: a discussion of theoretical implications. *J. Speech Hearing Res.*, 1971, *14*, 23-36.

Ades, A. Vowels, Consonants, Speech, and Nonspeech. *Psychol. Rev.*, 1977, *84*, 524-530.

Amcoff, S. Programmed instruction for Swedish children aged 7-10 years who are deaf or hard of hearing. *Am. Ann. Deaf*, 1968, *113*, 318-326.

Aslin, R. & Pisoni, D. Some developmental processes in speech perception. Paper presented at the NICHD conference 'Child Phonology: Perception, Production and Deviation', Maryland, 1978.

Baran, J., Zlatin, M. & Daniloff, R. Do mothers maximize phonetic contrastivity: a study of VOT. *J. Acoust. Soc. Am.*, 1976, *59*, 558 (A.).

Basso, A., Casati, G. & Vignolo, L. Phonemic identification defect in aphasia. *Cortex*, 1977, *13*, 85-95.

Beebe, H. Deaf children can learn to hear. *J. Comm. Dis.*, 1978, *11*, 193-200.

Bennet, C. & Ling, D. Effects of voiced-voiceless discrimination training upon articulation of hearing-impaired children. *Language and Speech*, 1977, *3*, 287-293.

Benton, A. Developmental aphasia and brain damage. *Cortex*, 1964, *1*, 40-52.

Benton, A. The cognitive functioning of children with developmental dysphasia. In *Developmental dysphasia*, M. Wyke ed., London: Academic Press, 1978, 43-62.

Bilger, R. & Wang, M. Consonant confusions in patients with sensorineural hearing loss. *J. Speech Hearing Res.*, 1976, *19*, 718-748.

Binnie, C., Montgomery, A. & Jackson, P. Auditory and visual contributions to the perception of consonants. *J. Speech Hearing Res.*, 1974, *17*, 619-630.

Blumstein, S., Baker, E. & Goodglass, H. Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, 1977b, *15*, 19-30.

Blumstein, S., Cooper, W., Zurif, E. & Caramazza, A. The perception and production of voice-onset-time in aphasia. *Neuropsychologia*, 1977a, *15*, 371-383.

Bocca, E.& Calearo, C. & Cassinari, V. A new method for testing hearing in temporal lobe tumors. *Acta Otolaryngol.*, 1954, *44*, 219-221.

Bocca, E., Calearo, C. Central hearing processes. In *Modern developments in audiology*,J. Jerger ed., N.Y.: Academic Press, 1963, 337-370.

Boothroyd, A. Speech perception and sensorineural hearing loss. In *Auditory management of hearing-impaired children*, M. Ross & Th. Giolas eds., Baltimore: University Park Press, 1978.

Bruner, J.S. On perceptual readiness. *Psychol. Rev.*, 1957, *64*, 123-152.

Bruner, J.S. Going beyond the information given. In *Beyond the information given*, J. Anglin ed., N.Y. Norton & Company, 1973, 218-238.

Burns, E. & Ward, W. Categorical perception of musical intervals. *J. Acoust. Soc. Am.*, 1974, *55* (A.).

Butterfield, E. & Cairns, G. Discussion summary-infant reception research. In *Language perspectives - acquisition, retardation, and intervention*, R. Schiefelbush & L. Lloyd eds., Macmillan, 1974, 75-102.

Calearo, C. & Antonelli, A. Cortical hearing tests and cerebral dominance. *Acta Otolaryngol.*, 1963, *56*, 17-26.

Calvert, D. Speech sound duration and the surd-sonant error. *Volta Review*, 1962, *64* , 401-402.

Campbell, H. *Phoneme recognition by ear and by eye*. Unpubl. doct. diss., Nijmegen: Stichting Studentenpers, 1974.

Capranica, R., Frishkopf, L. & Goldstein, M. Voice and hearing in the bullfrog. In *Models for the perception of speech and visual form* , W. Wathen-Dunn ed., Cambridge, Massachusetts: M.I.T. Press, 1965.

Carney, A., Widin, G. & Viemeister, N. Noncategorical perception of stop consonants differing in VOT. *J. Acoust. Soc. Am.*, 1977, *62*, 961-970.

Carpenter, R. & Rutherford, D. Acoustic cue discrimination in adult aphasia. *J. Speech Hearing Res.*, 1973, *16*, 534-544.

Cohen, J. & Diehl, F. Relation of speech-sound discrimination ability to articulation-type speech defects. *J. Speech Hearing Dis.*, 1963, *28*, 187-190.

Cole, R. & Scott, B. The phantom in the phoneme: Invariant cues for stop consonants. *Percept. Psychophys.*, 1974a, *15*, 101-107.

Cole, R. & Scott, B. Toward a theory of speech perception. *Psychol. Rev.*, 1974b, *81*, 348-374.

Cooper, W. Selective adaptation to speech. In *Cognitive Theory*, F. Restle, R. Shiffrin, N. Castellan, H. Lindman & D. Pisoni eds., N.Y.: John Wiley & Sons, 1975, Vol. I, 23-54.

Cooper, W., Billings, D. & Cole, R. Articulatory effects on speech perception: a second report. *J. Phonetics*, 1976, *4*, 219-232.

Cooper, W., Blumstein, S. & Nigro, G. Articulatory effects on speech perception: a preliminary report. *J. Phonetics*, 1975, *3*, 87-98.

Cooper, W. & Lauritsen, M. Feature processing in the perception and production of speech. *Nature*, 1974, *252*, 121-123.

Cooper, W. & Nager, R. Perceptuo-motor adaptation to speech: an analysis of bisyllabic utterances and a neural model. *J. Acoust. Soc. Am.*, 1975, *58*, 256-265.

Costello, J. Programmed instruction. *J. Speech Hearing Dis.*, 1977, *42*, 3-28.

Crul, Th. Een operante techniek voor auditieve discriminatietraining en het leren herkennen van geluiden. *Ned. tijdschr. psychol.*, 1971, *26*, 599-614.

Crul, Th. *Learning to perceive speech versus non-speech contrasts in hearing-disordered children: I. Auditory learning by programmed instruction*. Report MB. MED.PS. 2548, Kath. Univ. Nijmegen, the Netherlands, 1977a.

Crul, Th. *Learning to perceive speech versus non-speech contrasts in hearing-disordered children: II. The detection of covert perceptual, cognitive and semantic changes in auditory learning.* Report MB MED.PS. 2548, Kath. Univ. Nijmegen, the Netherlands, 1977b.

Crul, Th. & Peters, H. *Auditieve Discriminatie Test.* Handleiding, Amsterdam: Swets & Zeitlinger, 1976.

Cutting, J. A parallel between degree of encodedness and the ear advantage: Evidence from an ear-monitoring task. *J. Acoust. Soc. Am.*, 1973, *53*, 368 (A.).

Cutting, J. & Eimas, P. Phonetic feature analyzers and the processing of speech in infants. In *The role of speech in language*, J. Kavanagh & J. Cutting.eds., Cambridge, Massachusetts: M.I.T. Press, 1975, 127-148.

Danhauer, J. & Singh, S. *Multidimensional speech perception by the hearing-impaired.* Baltimore: University Park Press, 1975.

Darwin, C. The perception of speech. In *Handbook of perception*, E. Carterette & M. Friedman eds., N.Y.: Academic Press, 1976, 175-226.

Day, R. & Vigorito, J. A parallel between degree of encodedness and the ear advantage: Evidence from a temporal-order judgement task. *J. Acoust. Soc. Am.*, 1973, *53*, 368 (A.).

Documenta Geigy. *Scientific tables*, 5th edition. Basle: S. Karger, 1959.

Doehring, D. Picture-sound association in deaf children. *J. Speech Hearing Res.*, 1968, *11*, 49-62.

Doehring, D., Dudly, J. & Coderre, L. Programmed instruction in picture-sound association for the aphasic. *Folia Phoniat.*, 1967, *19*, 414-426.

Doehring, D. & Ling, D. Programmed instruction of hearing-impaired children in the auditory discrimination of vowels. *J. Speech Hearing Res.*, 1971, *14*, 746-754.

Eilers, R. & Minfie, F. Fricative discrimination in early infancy. *J. Speech Hearing Res.* 1975, *18*, 158-167.

Eilers, R., Wilson, W. & Moore, J. Developmental changes in speech discrimination in infants. *J. Speech Hearing Res.*, 1977, *20*, 766-780.

Eilers, R., Wilson, W. & Moore, J. Speech discrimination in the language-innocent and the language-wise: a study in the perception of voice onset time. *J. Child Lang.*, 1979, *6*, 1-18.

Eimas, P. The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, 1963, *6*, 206-217.

Eimas, P. Linguistic processing of speech by young infants. In *Language perspectives-acquisition, retardation, and intervention*, R. Schiefelbush & L. Lloyd eds., Macmillan, 1974, 55-73.

Eimas, P. Speech perception in early infancy. In *Infant perception: from sensation to cognition*, L. Cohen & Ph. Salapatec eds., N.Y.: Academic Press, 1975a, 193-231.

Eimas, P. Auditory and phonetic coding of the cues for speech: Discrimination and the /r-l/ distinction by young infants. *Percept. Psychophys.*, 1975b, *18*, 341-347.

Eimas, P., Cooper, W. & Corbit, J. Some properties of linguistic feature detectors. *Percept. Psychophys.*, 1973, *13*, 247-252.

Eimas, P. & Corbit, J. Selective adaptation of linguistic feature detectors. *Cognit. Psychol.*, 1973, *4*, 99-109.

Eimas, P., Siqueland, E., Jusczyk, P. & Vigorito, J. Speech perception in infants. *Science*, 1971, *171*, 303-306.

Eisenberg, R. *Auditory competence in early life.* The roots of communicative behaviour, Baltimore: University Park Press, 1976.

Erber, N. Speech-envelope cues as an acoustic aid to lipreading for profoundly deaf children. *J. Acoust. Soc. Am.*, 1972a, *51*, 1224-1227.

Erber, N. Auditory, visual and auditory-visual recognition of consonants by children with normal and impaired hearing. *J. Speech Hearing Res.*, 1972b, *15*, 413-422.

Ewing, I. & Ewing, A. *New opportunities for deaf children.* London: University of London Press, 1958.

Fairbanks, G. A theory of the speech mechanisms as a servosystem. *J. Speech Hearing Dis.*, 1954, *19*, 133-139.

Ferguson, G.A. *Statistical analysis in psychology and education.* N.Y.: McGraw-Hill Book Company, 1976.

Flower, R. Auditory disorders and reading disorders. In *Reading disorders, a multi-disciplinary symposium*, R. Flower ed., Philadelphia: F.A. Davis Company, 1965, 81-102.

Fodor, J., Garret, M. & Brill, S. Pi ka pu: The perception of speech sounds by prelinguistic infants. *Percept. Psychophys.*, 1975, *18*, 74-78.

Fourcin, A. Speech perception in the absence of speech productive ability. In *Language, cognitive deficits, and retardation.* N. O'Connor ed., London: Butterworths, 1975, 33-43.

Fourcin, A. Speech pattern tests for deaf children. *Speech and Hearing*, 1976, University College London, 47-62.

Fourcin, A. Acoustic patterns and speech acquisition. In *The development of communication*, N. Waterson & C. Snow eds., N.Y.: John Wiley & Sons, 1978.

Fourcin, A., Evershed, S., Fisher, J., King, A., Parker, A. & Wright, R. Perception and production of speech patterns by hearing-impaired children. *Speech and Hearing*, University college London, 1978, 174-204.

Friedlander, B. The effect of speaker identity, voice inflection, vocabulary, and message redundancy on infants selection of vocal reinforcement. *J. Exp. Child Psychol.*, 1968, *6*, 443-459.

Friedlander, B. Receptive language development in infancy. *Merrill-Palmer Quarterly*, 1970, *16*, 7-52.

Fromkin, V. Slips of the tongue. *Sci. Amer.*, 1973, *229*, 110-117.

Fry, D. The development of the phonological system in the normal and the deaf child. In *The genesis of language*, F. Smith & G. Miller eds., Cambridge, Massachusetts: M.I.T. Press, 1966, 187-206.

Fry, D. Phonological aspects of language acquisition in the hearing and the deaf. In *Foundations of language development*, E. Lenneberg & E. Lenneberg eds., N.Y.: Academic Press, 1975, 137-155.

Fry, D., Abramson, A., Eimas, P. & Liberman, A. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, *4*, 171-189.

Gemmil, J. & John, J. Time features in speech. *Teacher of the Deaf*, 1976, *74*, 386-402.

Gibson, E. *Principles of perceptual learning and development*. N.Y.: Appleton-Century Crofts, 1969.

Gilbert, H. & Campbell, M. Voice onset time in the speech of hearing-impaired individuals. *Folia Phoniat.*, 1978, *30*, 67-81.

Green, D. & Swets, J. *Signal detection theory and psychophysics*. N.Y.: John Wiley & Sons, 1966.

Griffiths, C. *Conquering Childhood Deafness*. New York: Exposition Press, 1967.

Guilford, J. *Fundamental statistics in psychology and education*. N.Y.: McGraw-Hill Book Company, 1965.

Haggard, M., Corrigall, J. & Legg, A. Perceptual factors in articulatory defects. *Folia Phoniat.*, 1971, *23*, 33-40.

Hart, J. 't & Cohen, A. Gating techniques as an aid in speech analysis. *Language and Speech*, 1964, *7*, 22-39.

Hartmann, H. Das wird schon noch! Zum Stande der Früherkennung und Früherförderung hörgeschädigter Kinder in der Bundesrepublik Deutschland. *Hörgeschädigter Kinder*, 1977, *14*, 10-12.

Hartmann, K. Zur Situation der Früherkennung und Früherförderung hörgeschädigter Kinder in der Bundesrepublik Deutschland. *Hörgeschädigter Kinder*, 1977, *14*, 6-9.

Heasly, B. *Auditory perceptual disorders and remediation*. Springfield: Charles C. Thomas, 1974.

Holland, A. & Matthews, J. Application of teaching machine concepts to speech pathology and audiology. *ASHA*, 1963, *5*, 474-482.

House, A., Stevens, K., Sandel, T. & Arnold, J. On the learning of speechlike vocabularies. *J. Verb. Learn. Verb. Beh.*, 1962, *1*, 133-143.

Ingram, T. Speech disorders in childhood. In *Foundations of language development*, Vol. 2, E. Lenneberg & E. Lenneberg eds., N.Y.: Academic Press, 1975, 195-261.

Irwin, R. Evaluating the perception of phonemes of children, ages 5 to 8. *J. Comm. Dis.*, 1974, *7*, 45-63.

Jakobson, R. *Child language and phonological universals*. The Haque: Mouton, 1968.

Jensema, C. & Trybus, R. *Communication patterns and educational achievement of hearing-impaired students*. Series T, number 2. Washington D.C.: Gallaudet College, Office of Demographic Studies, 1978.

Kent, R. Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. *J. Speech Hearing Res.*, 1976, *19*, 421-447.

Klatt, D. Voice onset time, frication, and aspiration in word-initial consonant clusters. *J. Speech Hearing Res.*, 1975, *18*, 686-706.

Kloster Jensen, M. & Jussen, H. *Lautbildung bei Hörgeschädigten*. Berlin: Marhold, 1970.

Korsan-Bengtsen, M. Distorted speech audiometry. *Acta Otolaryngol*, suppl. 310, 1973.

Kruger, F., Stromberg, H. & Levitt, H. Synthetic speech as a diagnostic tool. *Communication Sciences Laboratory Research Report number 2*, 1972, City University of New York Graduate School.

Kuhl, P. & Miller, D. Speech perception by the Chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, *190*, 69-72.

Kuhl, P. & Miller, D. Speech perception by the Chinchilla: Identification functions for synthetic VOT-stimuli. *J. Acoust. Soc. Am.*, 1978, *63*, 905-916.

Lane, H. The motor theory of speech perception: a critical review. *Psychol. Rev.*, 1965, *72*, 275-309.

Lane, H. & Moore, D. Reconditioning a consonant discrimination in an aphasic: an experimental case history. *J. Speech Hearing Dis.*, 1962, *27*, 232-243.

Lehiste, I. The units of speech perception. In *Speech and corticál functioning*, J. Gilbert ed., N.Y.: Academic Press, 1972, 187-235.

Lenneberg, E. Understanding language without ability to speak: A case report. *J. abnorm. soc. Psychol.*, 1962, *65*, 419-425.

Liberman, A. Some results of research on speech perception. *J. Acoust. Soc. Am.*, 1957, *29*, 117-123.

Liberman, A. The grammars of speech and language. *Cognit. Psychol.*, 1970, *1*, 301-323.

Liberman, A. The speech code. In *Communication, language and meaning* , G. Miller ed., N.Y.: Basic Books Inc., 1973a, 128-140.

Liberman, A. The specialization of the language hemisphere. In *The Neurosciences: Third Study Program*, F. Schnitt & F. Worden eds., Cambridge: M.I.T. Press, 1973b, 43-56.

Liberman, A., Cooper, F., Harris, K. MacNeilage, P. & Studdert-Kennedy, M. Some observations on a model for speech perception. In *Models for the perception of speech and visual form*, W. Wathen-Dunn ed., Cambridge, Massachusetts: M.I.T. Press, 1967a, 68-87.

Liberman, A., Cooper, F., Shankweiler, D. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.*, 1967b, *74*, 431-461.

Liberman, A., Delattre, P. & Cooper, F. The rôle of selected stimulus-variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.*, 1952, *65*, 497-516.

Liberman, A., Delattre, P. & Cooper, F. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1958, *1*, 153-167.

Liberman, A., Harris, K., Hoffman, H. & Griffith, B. The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.*, 1957, *54*, 358-568.

Liberman, A., Harris, K. Kinney, J. & Lane, H. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. exp. Psychol.*, 1961a, *61*, 379-388.

Liberman, A., Harris, K., Eimas, P., Lisker, L. & Bastian, J. An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 1961b, *4*, 175-195.

Lidén, G. Speech audiometry. An experimental and clinical study with Swedish language material. *Acta Otolaryngol.* suppl. 114, 1954.

Ling, D. Auditory coding and recoding: An analysis of auditory training procedures for hearing-impaired children. In *Auditory management of hearing-impaired children,* M. Ross & Th. Giolas eds., Baltimore: University Park Press, 1978.

Ling, D. & Doehring, D. Learning limits of deaf children for coded speech. *J. Speech Hearing Res.*, 1969, *12*, 83-94.

Ling, D. & Ling, A. *Aural habilitation:* the foundations of verbal learning in hearing-impaired children. Washington: The Alexander Graham Bell Association for the Deaf, 1978.

Lisker, L. On learning a new contrast. *Stat. Rep. Hask. Lab., AD-727 616,* 1970.

Lisker, L. In qualified defense of VOT. *Language and Speech*, 1978, *21*, 375-383.

Lisker, L. & Abramson, A. A cross-language study of voicing initial stops: acoustical measurements. *Word*, 1964, *20*, 384-419.

Lisker, L. & Abramson, A. Some effects of context on voice onset time in English stops. *Language and Speech*, 1967, *10*, 1-28.

Lisker, L., Cooper, F. & Liberman, A. The uses of experiment in language description. *Word*, 1962, *18*, 82-106.

Lisker, L., Liberman, A., Erickson, D., Dechovitz, D. & Mandler, R. On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, 1977, *20*, 209-216.

Locke, S. & Keller, L. Categorical perception in a non-linguistic mode. *Cortex*, 1973, *9*, 355-369.

Löwe, A. *Früherfassung, Früherkennung, Frühbetreuung hörgeschädigter Kinder.* Berlin: Marhold, 1970.

Löwe, A. Die Früherkennung, Früherfassung und Früherförderung hörgeschädigter Kinder aus pädagogischer Sicht. *Hörgeschädigter Kinder*, 1977, *14*, 14-21.

Lowe, A. & Campbell, R. Temporal discrimination in aphasoid and normal children. *J. Speech Hearing Res.*, 1965, *8*, 313-314.

MacNeilage, P., Rootes, Th. & Chase, R. Speech production and perception in a patient with severe impairment of somesthetic and motor control. *J. Speech Hearing Res.*, 1967, *10*, 450-467.

Mark, H. & Hardy, W. Orienting reflex disturbances in central auditory or language handicapped children. *J. Speech Hearing Dis.*, 1958, *23*, 237-242.

Mártony, J. & Agelfors, E. Two psychoacoustic tests with severely hard of hearing children. In *Speech Communication*, G. Fant ed., N.Y.: John Wiley & Sons, 1975, 69-76.

Maspétiol, R. & Semette, D. Les test d'atteinte auditive corticale et centrale. *Acta Otolaryngol.*, 1964, *58*, 459-470.

Massaro, D. *Understanding language*. N.Y.: Academic Press, 1975.

Mattingly, I., Liberman, A., Syrdal, A. & Halwes, T. Discrimination in speech and nonspeech mode. *Cognit. Psychol.*, 1971, *2*, 131-157.

McReynolds, L. Operant conditioning for investigating speech sound discrimination in aphasic children. *J. Speech Hearing Res.*, 1966, *9*, 519-528.

McReynolds, L. & Engmann, D. *Distinctive feature analysis of misarticulations*. Baltimore: University Park Press, 1975.

Miller, C. & Morse, Ph. The 'heart' of categorical speech discrimination in young infants. *J. Speech Hearing Res.*, 1976, *19*, 578-589.

Miller, J. Nonindependence of feature processing in initial consonants. *J. Speech Hearing Res.*, 1977, *20*, 519-528.

Miller, G. & Nicely, P. An analysis of perceptual confusions among some english consonants. *J. Acoust. Soc. Am.*, 1955, *27*, 338-352.

Miyawaki, K., Liberman, A., Fuyimura, O., Strange, W. & Jenkins, J. Cross-language study of the perception of the F_3 cue for /r/ versus /l/ in speech and nonspeech-like patterns. In *Auditory analysis and perception of speech*, G. Fant & M. Tatham eds., London: Academic Press, 1975b, 339-348.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J. & Fuyimura, O. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Percept. Psychophys.*, 1975a, *18*, 331-340.

Moffit, A. Consonant cue perception by twenty- to twenty-four week old infants. *Child Develop.*, 1971, *42*, 717-731.

Monnin, L. & Huntington, D. Relationship of articulatory defects to speech sound identification. *J. Speech Hearing Res.*, 1974, *17*, 352-366.

Morse, P. The discrimination of speech and nonspeech stimuli in early infancy. *J. exp. Child Psychol.*, 1972, *14*, 477-492.

Morse, Ph. Infant speech perception: a preliminary model and review of the literature. In *Language perspectives - acquisition, retardation, and intervention*, R. Schiefelbusch & L. Lloyd eds., Macmillan, 1974, 19-53.

Murphy, K. Auditory training. Research dilemmas. *Teacher of the Deaf*, 1979, 3, 46-54.

Nooteboom, S. & Cohen, A. *Spreken en verstaan*: Een inleiding tot de experimentele fonetiek. Assen: van Gorcum, 1976.

Oakland, Th. & Williams, F. *Auditory perception*. Seatlle: Special Child Publications, 1971.

Ostle, B. *Statistics in Research*, Iowa State University Press, 1963.

Owens, E. Consonant errors and remediation in sensorineural hearing loss. *J. Speech Hearing Dis.*, 1978, *43*, 331-347.

Owens, E., Benedict, M. & Schubert, E. Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairment. *J. Speech Hearing Res.*, 1972, *15*, 308-322.

Palva, A. Filtered speech audiometry. *Acta Otolaryngol.*, suppl. 210, 1965.

Pickett, J. & Mártony, J. Low frequency vowel formant discrimination in hearing-impaired listeners. *J. Speech Hearing Res.*, 1970, *13*, 347-359.

Pisoni, D. & Sawusch, J. Some stages of processing in speech perception. In *Structure and process in speech perception*, A. Cohen & S. Nooteboom eds., Berlin: Springer-Verlag, 1975, 16-34.

Pollack, D. *Educational audiology for the limited hearing infant*. Springfield: Charles C. Thomas, 1970.

Pollack, I. & Norman, D. A non-parametric analysis of recognition experiments. *Psychon. Sci.*, 1964, *1*, 125-126.

Pollack, I. & Pisoni, D. On the comparison between identification and discrimination tests in speech perception. *Psychon. Sci.*, 1971, *24*, 299-300.

Pols, L. & Schouten, M. Identification of deleted consonants. *J. Acoust. Soc. Am.*, 1978, *64*, 1333-1337.

Potter, R., Kopp, G. & Green, H. *Visible Speech*. N.Y.: Dover Publications, 1966.

Pronovost, W. & Dumbleton, C. A picture-type speech sound discrimination test. *J. Speech Hearing Dis.*, 1953, *18*, 258-266.

Rees, N. Auditory processing factors in language disorders: the view from Procrustes' bed. *J. Speech Hearing Dis.*, 1973, *38*, 304-314.

Saleh, H. Sights and sounds: An auditory program for young deaf children. *Am. Ann. Deaf.*, 1965, *110*, 528-534.

Schulte, K. Zum Einsatz von Vibrations-Verstärkern als Sprechgliederungs-Hilfe. Stellungnahmen zu Fragen aus der Praxis. *Hörpäd.*, 1979, *33*, 95-101.

Schultz, M. & Kraat, A. Lack of perceptual reality of the phoneme for hearing handicapped children. *Language and Speech*, 1971, *14*, 178-186.

Shankweiler, D., Strange, W. & Verbrugge, R. Speech and the problem of perceptual constancy. *Status Report SR - 42/43*, Haskins Lab., 1975, 117-145.

Shankweiler, D. & Studdert-Kennedy, M. Identification of consonants and vowels presented to left and right ears. *Quart. J. exp. Psychol.*, 1967, *19*, 59-63.

Shvachkin, N. The development of phonemic speech perception in early childhood. In *Studies of child development*, C. Ferguson & D. Slobin eds., N.Y.: Holt, Rinehart & Winston, 1973, 91-127.

Siegel, S. *Nonparametric statistics*. N.Y.: McGraw-Hill, 1956.

Silverman, F. *Research design in speech pathology and audiology*, New Jersey: Prentice-Hall, 1977.

Simon, C. Some aspects of the development of speech production and perception in children. In *Speech Communication*, vol. 4, G. Fant ed., N.Y.: Halsted Press, 1975.

Simon, C. On the use of comfortable listening levels in speech experiments. *J. Acoust. Soc. Am.*, 1978, *64*, 744-750.

Simon, C. & Fourcin, A. Cross-language study of speech pattern hearing. *J. Acoust. Soc. Am.*, 1978, *63*, 925-935.

Singh, S. Distinctive Features: A measure of consonant perception. In *Measurement procedures in speech, hearing, and language*, S. Singh ed., Baltimore: University Park Press, 1975.

Slis, I. & Cohen, A. On the complex regulating the voiced-voiceless distinction I. *Language and Speech*, 1969a, *12*, 80-102.

Slis, I. & Cohen, A. On the complex regulating the voiced-voiceless distinction II. *Language and Speech*, 1969b, *12*, 137-155.

Spiegel, M. *Theory and problems of statistics*, N.Y.: McGraw-Hill, 1961.

Spriesterbach, D. & Curtis, J. Misarticulation and discrimination of speech sounds. In *Articulation testing and treatment*, E. McDonald ed., Pittsburgh: Stanwix House, 1964, 41-54.

Stark, R. *Sensory capabilities of hearing-impaired children*. Baltimore: University Park Press, 1974.

Stevens, K. Segments, features and analysis by synthesis. In *Language by Ear and by Eye*, J. Kavanagh & I. Mattingly eds., Cambridge, Massachusetts: M.I.T. Press, 1972, 47-52.

Stevens, K. The potential role of property detectors in the perception of consonants. In *Auditory analysis and perception of speech*, G. Fant & M. Tatham eds., London: Academic Press, 1975, 303-330.

Stevens, K. & House, A. Speech Perception. In *Foundations of modern auditory theory*, J. Tobias ed., N.Y.: Academic Press, 1972, 1-57.

Stevens, K. & Klatt, D. The role of formant transitions in the voiced-voiceless distinction for stops. *J. Acoust. Soc. Am.*, 1974, *55*, 653-659.

Stevens, K., Liberman, A., Studdert-Kennedy, M. & Öhman, S. Cross-language study of vowel perception. *Language and Speech*, 1969, *12*, 1-23.

Stewart, J., Singh, S. & Hayden, M. Distinctive feature use in speech perception of children. *Language and Speech*, 1979, *22*, 69-79.

Streeter, L. Language perception of two-month old infants shows effects of both innate mechanisms and experience. *Nature*, 1976, *259*, 39-41.

Studdert-Kennedy, M. The perception of speech. *Status Report AD 723 586*, Haskins Lab., New Haven, Connecticut, 1970, 15-48.

Studdert-Kennedy, M. From acoustic signal to phonetic message. *J. Comm. Dis.*, 1975, *8*, 181-188.

Studdert-Kennedy, M., Liberman, A., Harris, K. & Cooper, F. Motor theory of speech perception: a reply to Lane's critical review. *Psychol. Rev.*, 1970, *77*, 234-249.

Studdert-Kennedy, M. & Shankweiler, D. Hemispheric specialization for speech perception. *J. Acoust. Soc. Am.*, 1970, *48*, 579-594.

Tallal, P. & Newcombe, F. What can computer-synthesized speech tell us about the language comprehension impairment of adults with residual dysphasia? *J. Acoust. Soc. Am.*, 1976, *59*, suppl. 1, 585 (A.).

Tallal, P. & Piercy, M. Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia*,1973, *11*, 389-398.

Tallal, P. & Piercy, M. Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*, 1974, *12*, 83-93.

Tallal, P. & Piercy, M. Developmental aphasia: The perception of brief vowels and extented stop consonants. *Neuropsychologia*, 1975, *13*, 69-74.

Tallal, P. & Piercy, M. Defects of auditory perception in children with developmental dysphasia. In *Developmental dysphasia*, M. Wyke ed., London: Academic Press, 1978, 63-84.

Templin, M. *Certain language skills in children.* Minneapolis: University of Minnesota Press, 1957.

Travis, L. & Rasmus, B. The speech sound discrimination ability of cases with functional disorders of articulation. *Quart. J. Speech*, 1931, *17*, 217-226.

Trehub, S. & Rabinovitch, M. Auditory-linguistic sensitivity in early infancy. *Develop. Psychol.*, 1972, *6*, 74-77.

Uden, A. van *Dove kinderen leren spreken.* Doct. Dissert., Rotterdam: Universitaire Pers, 1974.

Uselding, D. A temporal order effect in voice onset time discrimination. *Language and Speech*, 1977, *20*, 366-376.

Van Riper, Ch. & Irwin, J. *Voice and articulation*, N.Y.: Prentice Hall, 1958.

Wang, M., Reed, C. & Bilger, R. A comparison of the effects of filtering and sensorineural hearing loss on patterns of consonant confusions. *J. Speech Hearing Res.*, 1978, *21*, 5-36.

Waters, R. & Wilson, W. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Percept. Psychophys.*, 1976, *4*, 285-289.

Watts, W. Auditory training. *Teacher of the Deaf*, 1969, *67*, 4-18.

Wedenberg, E. Auditory training of deaf and hard of hearing children. *Acta Otolaryngol.*, suppl. 94, 1951.

Wedenberg, E. Auditory training of severely hard of hearing pre-school children. *Acta Otolaryngol.*, suppl. 110, 1954.

Weiner, P. Auditory discrimination and articulation. *J. Speech Hearing Dis.*, 1967, *32* , 19-27.

Wepman, J. *Auditory discrimination test.* Manual, Chicago: Lang. Res. Ass., 1958.

Whetnall, E. & Fry, D. *The deaf child.* London: Heineman, 1971.

Wickelgrenn, W. Distinctive features and errors in short-term memory for English vowels. *J. Acoust. Soc. Am.*, 1965, *38*, 583-588.

Wickelgrenn, W. Distinctive features and errors in short-term memory for English consonants. *J. Acoust. Soc. Am.*, 1966, *39*, 388-389.

Winitz, H. *Articulatory acquisition and behaviour.* N.Y.: Appleton, 1969.

Winitz, H. & Bellerose, B. Effects of pretraining on sound discrimination learning. *J. Speech Hearing Res.*, 1963, *6*, 171-180.

Winitz, H. & Preisler, L. Discrimination pretraining and sound learning. *Percept. Mot. Skills.*, 1965, *20*, 905-916.

Zigmund, N. & Cicci, R. *Auditory learning.* Belmont: Dimensions Publishing Co., 1968.

Zlatin, M. Voicing contrast: perceptual and productive voice onset time characteristics of adults. *J. Acoust. Soc. Am.*, 1974, *56*, 981-994.

Zlatin, M. & Koenigsknecht, R. Development of the voicing contrast perception of stop consonants. *J. Speech Hearing Res.*, 1975, *18*, 530-540.

Zlatin, M. & Koenigsknecht, R. Development of the voicing contrast: a comparison of voice onset time in stop perception and production. *J. Speech Hearing Res.*, 1976, *19*, 93-111.

Appendix A

/ t - d /lijst

1	[dɑm	tɑm]
2	[di:rən	ti:rən]
3	[tɑ͜ᵘʋ	dɑ͜ᵘʋ]
4	[dɪk	tɪk]
5	[tœ͜ʏt	dœ͜ʏt]
6	[ta:kən	da:kən]
7	[do͜:ɪ	to͜:ɪ]
8	[tɑl	dɑl]
9	[de:lən	te:lən]
10	[tu:	du:]
11	[de:xən	te:xən]
12	[tɑk	dɑk]
13	[te:rən	de:rən]
14	[di:n	ti:n]
15	[da:lən	ta:lən]
16	[dɛŋk	tɛŋk]
17	[tɔr	dɔr]
18	[do:s	to:s]
19	[tʏ:rən	dʏ:rən]
20	[de:kən	te:kən]
21	[tɔp	dɔp]
22	[tɑs	dɑs]
23	[do:rən	to:rən]
24	[dɔrsən	tɔrsən]
25	[tœrf	dœrf]
26	[dœ͜ʏn	tœ͜ʏn]

Appendix B / p - b /lijst

1	[pʏːr	bʏːr]	22	[pɑs	bɑs]
2	[baːrt	paːrt]	23	[bɑk	pɑk]
3	[beːn	peːn]	24	[pɛt	bɛt]
4	[pøl	bøl]	25	[boːs	poːs]
5	[pɑrt	bɑrt]	26	[puf	buf]
6	[pɛn	bɛn]	27	[biːt	piːt]
7	[bɔl	pɔl]	28	[pul	bul]
8	[piːr	biːr]	29	[bɑt	pɑt]
9	[bɔf	pɔf]	30	[pɛk	bɛk]
10	[paːs	baːs]	31	[bœl	pœl]
11	[p$\underset{\smile}{\text{ɑ}}^{\text{u}}$ʋ	b$\underset{\smile}{\text{ɑ}}^{\text{u}}$ʋ]	32	[poːt	boːt]
12	[beːr	peːr]	33	[bɪkən	pɪkən]
13	[baːl	paːl]	34	[bɛrk	pɛrk]
14	[pɛ$\underset{\smile}{\text{s}}$t	bɛ$\underset{\smile}{\text{s}}$t]	35	[bøk	pøk]
15	[pɑn	bɑn]	36	[pɑl	bɑl]
16	[bɛlən	pɛlən]	37	[bɔnt	pɔnt]
17	[poːrt	boːrt]	38	[pɔt	bɔt]
18	[bɪŋk	pɪŋk]	39	[pɪt	bɪt]
19	[p$\underset{\smile}{\text{œ}}^{\text{i}}$	b$\underset{\smile}{\text{œ}}^{\text{i}}$]	40	[bɑnt	pɑnt]
20	[b$\underset{\smile}{\text{ɛ}}^{\text{ɪ}}$l	p$\underset{\smile}{\text{ɛ}}^{\text{ɪ}}$l]	41	[pɔk	bɔk]
21	[baːr	paːr]	42	[bœk	pœk]

Appendix C / t - d - b - p /lijst

1 [biːr tiːr piːr diːr]

2 [tɑ͜ᵘɒ bɑ͜ᵘɒ pɑ͜ᵘɒ dɑ͜ᵘɒ]

3 [paːlən daːlən baːlən taːlən]

4 [doːs poːs toːs boːs]

5 [bɔt pɔt dɔt tɔt]

6 [deːn beːn peːn teːn]

7 [bɑk tɑk pɑk dɑk]

8 [pɑs dɑs bɑs tɑs]

9 [dʏːrən pʏːrən bʏːrən tʏːrən]

10 [pɔp tɔp bɔp dɔp]

11 [tɔf bɔf dɔf pɔf]

12 [pɔl dɔl bɔl tɔl]

13 [pɑl bɑl dɑl tɑl]

14 [pɪn dɪn bɪn tɪn]

15 [tɛl bɛl pɛl dɛl]

16 [bɪk tɪk pɪk dɪk]

APPENDIX D

Tchupproff's coefficient of contingency.

It had been established in section 5.2.3.2. that there were significant differences between the predicted and observed confusion matrices both before and after training. This pointed to the fact that, statistically, there was no complete correspondence between them. From a statistical point of view this meant that the voice features and place features were not perceived completely independent of each other. The X^2 values, however, did not show how closely these features were related to each other. As the two variables might be related in a significant way, whereas at the same time the intensity of the association could be rather low, we looked for a coefficient which would enable us to determine the degree of association between the [place] and [voice] variables. For this purpose we used Tchupproff's coefficient of contingency (cf. Gadourek, *A Dutch Community*. Leiden, Stenfert Kroese, 1956, 307-308) by applying the formula:

$$T = \sqrt{\frac{\frac{X^2}{N}}{\sqrt{(r-1)(k-1)}}}$$

where N means the total number of observations (here 1536), r the number of rows (4) and k the number of columns (4) of the matrices.
T covers the values from 0 to 1, being 0 if $X^2=0$ and 1 if

$$\sqrt{\frac{\frac{X^2}{N}}{\sqrt{(r-1)(k-1)}}} = 1 \qquad \text{or} \qquad X^2 = N\sqrt{(r-1)(k-1)}$$

In our research the X^2-values of the pre-training and post-training matrices had to amount to the value of $1536\sqrt{9} = 4608$ in order to get a complete association of T=1. This value is only hypothetical and extremely high in comparison with the observed $X^2 = 52,75$ of the pre-training matrix and the $X^2 = 72,17$ of the post-training matrix. Correspondingly, the respective T-values are relatively low, viz. T=.11 for the pre-training matrix and T=.13 for the post-training matrix. It appears, therefore, that though the association between place features and voice features is significant according to the X^2-test, it is only of negligible intensity.

CURRICULUM VITAE

Thom Crul werd op 9 augustus 1935 te Sawah Loento op het eiland Sumatra geboren. Na het behalen van het gymnasium-alpha diploma te Breda volbracht hij de militaire dienstplicht en vervolgens studeerde hij psychologie aan de Katholieke Universiteit van Nijmegen. Hier legde hij het doktoraal-eksamen af in de hoofdrichting ontwikkelingspsychologie (Prof.Dr. P.J.A. Calon). Hij trad in dienst van de Afdeling Medische Psychologie (Prof.Dr. P.B. Bierkens) van het St. Radboudziekenhuis en werd als wetenschappelijk medewerker verbonden aan de Afdeling Kinderaudiologie, Keel-, neus- en oorheelkunde (Prof.Dr. W.F.B. Brinkman). Vanaf het begin van zijn werkzaamheden ging zijn belangstelling uit naar de kognitieve ontwikkelingsmogelijkheden van het kommunikatief-gestoorde kind. Hij richtte zich met onderzoek vooral op de edukatief-audiologische aspekten van jonge gehoorgestoorde kinderen. In dit verband verschenen van hem enkele wetenschappelijke publikaties en in ko-auteurschap met een kollega van de Afdeling Medische Psychologie (Drs. H.F.M. Peters) een Auditieve Diskriminatietest.

Pas nadat aan de Nijmeegse Universiteit de Interfakultaire Werkgroep Taal- en Spraakgedrag werd opgericht (Dr. J.C. Marshall), waarbinnen hij aan enkele projekten participeert, en na het volgen van een inleiding in de Fonetiek op het Fonetisch Laboratorium van de Universiteit van Nijmegen (Prof.Dr. W.H. Vieregge), werd zijn interesse speciaal gericht op de spraakwaarnemingsproblemen van het auditief gehandicapte kind. Deze interesse leidde tot het onderzoek dat ten grondslag ligt aan dit proefschrift en dat mede door faciliteiten op de Afdeling Funktieleer van de Psychologische Subfakulteit te Nijmegen (Prof.Dr. W.J.M. Levelt) kon worden gekonkretiseerd.

STELLINGEN

1. Auditieve training in de vorm van het leren diskrimineren tussen twee kontrasterende distinktieve kenmerken binnen een beperkte kontekst, heeft een generaliserend effekt op de spraakwaarneming van het gehoorgestoorde kind. (dit proefschrift).

2. Het auditief leren diskrimineren van het linguistisch relevante verschil tussen stemhebbende- en stemloze beginmedeklinkers kan worden verklaard op grond van de aktiviteit van een auditief dekoderingsmechanisme en vereist niet noodzakelijk de bemiddeling van processen die bij het spreken een rol spelen. (dit proefschrift).

3. Het verdient aanbeveling om kinderen die bij het uitvoeren van leesvoorwaarde-toetsen als klankanalyse, klanksynthese en auditieve diskriminatie problemen hebben, met behulp van een kategorische spraakperceptie-test verder te onderzoeken op hun bekwaamheid om geïsoleerde spraakgeluiden te identificeren.
(Liberman, I.; Shankweiler, D.; Fisher, F. & Carter, B., Explicit syllable and phoneme segmentation in the young child. *J. Exp. Child Psychol.*, 1974, *18*, 201-212.
Liberman, I.; Shankweiler, D.; Liberman, A.; Fowler, C. & Fisher, F., Phonetic segmentation and recoding in the beginning reader. In: *Toward a psychology of reading*, Reber, A. & Scarborough, D. (eds.), N.Y.: John Wiley & Sons, 1977).

4. De tussenpositie die de onomatopee niet alleen inneemt binnen de semantische hiërarchie van konkreet-betekenisvol geluid tot abstrakt woord, maar ook ten opzichte van de essentieel verschillende akoestische patronen van een niet-spraakgeluid en gekodeerde spraak, maakt dat het een makkelijk toepasbaar en funktioneel middel is voor het aanleren van primair receptief- en expressief taalgedrag bij jonge hoorgestoorde kinderen.

5. In tegenstelling tot hun traditionele naam zijn de fonetisch gebalanceerde spraaklijsten, waarvan bij de klinische audiometrie gebruik wordt gemaakt, niet fonetisch- maar fonemisch gebalanceerd.

6. Testwoorden van spraakbanden die bestemd zijn voor audiologisch onderzoek en die geïsoleerd of in citatie-vorm werden uitgesproken, zijn wat betreft hun kritisch akoestische informatie onnatuurlijk redundant in vergelijking met woorden die uit lopende spraak zijn geselekteerd. De verstaanbaarheid van deze woorden dient mede in dit licht te worden geïnterpreteerd. (Lisker, L. & Abramson, A. Some effects of context on voice onset time in English stops. *Language and Speech,* 1967, *10,* 1-28).

7. Bij het ontwerpen en aanpassen van hoortoestellen wordt ten onrechte meer aandacht geschonken aan geluidsversterking van de spraakfrekwenties dan aan een versterking van de frekwenties die het mogelijk maken om relevante omgevingsgeluiden waar te nemen.

8. In verband met de signifikant slechtere fundamentele luisterkapaciteiten voor spraak die slechthorende kinderen 's middags op school vertonen in vergelijking met 's ochtends, moet er naar worden gestreefd om het lesrooster zo in te delen dat alle belangrijke verbale lessen op een school voor slechthorenden gedurende de ochtenduren worden gegeven.

9. Bij de traditionele vertoning van oude ambachten tijdens gevarieerde evenementen in Nederland ontbreekt stelselmatig het oudste.

Th.A.M. Crul Malden, 1980.