Chapter 1

General introduction

1.1 Introduction

One of the processes that contribute to the comprehension of speech is the segmentation of the speech stream into separate phonemes (vowels and consonants). Without this segmentation process, the speech signal is a meaningless blur of sounds, as it is when one listens to an unfamiliar foreign language.

Although the segmentation process is important, it is not always necessary to categorise each sound in order to comprehend a spoken message. Listeners are very flexible in their use of meaningful speech units, i.e. phrases, words, syllables, and phonemes. What units the listener will use depends largely on the 'noise' in the acoustic signal and the 'predictability' of the message. In the case of stock phrases like "How are you?" or "Have a nice day", listeners need relatively little information about the individual segments for recognition, because these highly frequent utterances are stored as a whole in long-term memory. In addition, they are associated with a very specific speech situation: "How are you?" belongs to the situation in which one person meets another person. Therefore, the speech situation will predict this utterance to a high degree, which also reduces the need to actually process all individual segments.

With less predictable speech utterances, the listener needs to analyse the acoustic speech signal in more detail, e.g. analyse the utterance into separate words. However, in many cases environmental noise or sloppy articulation will cause the speech stream to be partly incomprehensible. Despite these 'gaps' in the speech signal, the listener will usually not have much difficulty comprehending the message, because the semantic and syntactic context will predict to a very great extent the words that are distorted. For instance, in the sentence "John

was writing a letter to his ….", the missing word is more likely to be "girlfriend", than "dog" or "Wednesday".

The most demanding speech comprehension situation for the listener is speech with low predictability, such as words produced in isolation, proper names, or unfamiliar words. In this situation the listeners will need to analyse the speech signal into successive speech segments. This task is not as easy as it may seem, because the acoustic speech signal is characterised by high variability. Even in the realisation of *identical* speech sounds (e.g. a repetition of the word "doll") there is a large variation due to talker-specific factors, such as gender, individual voice characteristics, speech rate, dialect, etc.

Apparently, the listener has *learned* to abstract away from the acoustic input and to use only the information that distinguishes one speech sound from another, but to ignore sometimes widely varying acoustic properties that do not distinguish speech sounds in the language. This abstraction from the acoustic signal towards phonemes is called categorical perception (categorical perception is not a special speech mechanism, but a common cognitive principle).

1.2 Categorical speech perception

In the past 40 years of speech perception research there has been a special interest in categorical perception, especially in the ability or inability of listeners to perceive differences between speech sounds that belong to the same phoneme category. In general it has been claimed that listeners have difficulty perceiving differences between different realisations of the *same* phoneme, but that they find it easy to hear differences between *different* phoneme categories. As a result of decades of research, 'categorical perception' has become identified with a particular laboratory paradigm and with particular results. In the 'categorical perception paradigm', a stimulus continuum between speech sounds is presented to listeners in two psychoacoustic tasks: a discrimination task and a classification task. 'Categorical perception results' are results that demonstrate a strong relationship between listeners' discrimination and classification performance: speech stimuli classified as belonging to the same category are difficult to discriminate, whereas stimuli classified as belonging to different categories are easy to discriminate. This means that discrimination performance is predictable from classification performance.

In the first categorical perception study, Liberman, Harris, Hoffman, and Griffith (1957) investigated listeners' classification and discrimination of a stimulus continuum between the stop consonants /b-d-γ/. Liberman et al. (1957) concluded that their results did not agree with their own "extreme assumption": discrimination results were better than predicted by classification. This outcome apparently represented the listener's ability to distinguish the speech sounds not solely on the basis of the phonemic labels, but also on the basis of the acoustic differences between the stimuli. In spite of the authors' conclusion, however, this first study is often cited as a paradigmatic example of categorical perception (for a review, see Repp, 1984). Even though a strict relationship between discrimination and classification has rarely been demonstrated in subsequent research, results are often interpreted in terms of absolutely "categorical" or "continuous" perception (Macmillan, 1987).

In continuous perception, discrimination performance is not restricted by phoneme categorisation, but based on the acoustic differences between the stimuli. As a result, discrimination is unrelated to classification performance, and typically much better than predicted by classification (Massaro & Cohen, 1983). However, discrimination results that are interpreted as indicating categorical perception are often also much better than predicted by

classification. In other words, it is not clear on what basis the distinction between categorical and continuous perception can be made. There is no explicit criterion for the maximum difference between discrimination and classification results that would still be compatible with 'categoricalness'. This has consequences for the interpretation of differences in the relationship between classification and discrimination performance, for instance between stop consonants and vowels. Stop consonants are said to be categorically perceived, whereas the perception of vowels is continuous (Fry et al., 1962; Pisoni, 1973; Repp, 1981; Stevens et al., 1969). In view of the lack of a clear distinction between categorical and continuous perception it would be better to interpret the results as being *more* categorical for stop consonants than for vowels, rather than making an arbitrary binary distinction between categorical and continuous perception (Studdert-Kennedy, et al., 1970). We use the original definition of Liberman et al. (1957): perception is fully categorical only if there is no significant difference between phoneme categorisation and discrimination.

Despite the unclear criterion for categorical perception there have been numerous demonstrations of a fairly strong relationship between discrimination and classification of speech sounds. But what is it that makes listeners' classification performance predict their discrimination performance? How are the categorical perception results explained?

1.3 Explaining the categorical perception results

Since phoneme categories are language-specific rather than universal, it seems obvious that the categorical perception data are best explained with reference to extensive experience with the native language (Fujisaki & Kawashima, 1970, 1971; Moore, 1997; Pisoni, 1973, 1975; Rosen & Howell, 1987). When we learn the phonological system of a particular language, we learn to attend to acoustic differences which affect the meanings of words, and to ignore acoustic differences which do not affect word meanings. Once we have mastered this, it may be difficult to hear acoustic differences which do not affect word meanings. A strong relationship between discrimination and classification of speech sounds will arise as a natural consequence of this.

A different explanation for the categorical perception results came from Liberman and colleagues, who claimed that they formed evidence for a unique speech mechanism, and hence that categorical perception should not occur for nonspeech sounds (e.g. Liberman et al. 1957; Liberman et al., 1961, or see the collected papers in Liberman, 1996). The basic idea behind this view, represented in the Motor Theory of Speech Perception, is that perception is strongly influenced by speech production. Discontinuities in speech perception, e.g. the phoneme boundaries between /pa-ta/, are supposed to reflect discontinuities in the production of these speech sounds, such as the discontinuity between a movement of the lips for /pa/ and a movement of the tongue for /ta/. However, the motor theory could not explain some of the highly categorical results with speech sounds, such as the voicing contrast /do-to/, for which it is difficult to show that there is a strong discontinuity in production (Liberman et al., 1961). Nor could it explain the relatively low degree of categorical perception for the voiceless affricate/fricative contrast /τΣA-ΣA/, for which there is a clear discontinuity in articulation (Howell & Rosen, 1984). Interestingly, in more recent versions of the Motor Theory of Speech Perception, categorical perception has been abandoned as an argument in support of the correlation between speech production and perception (Liberman, 1996).

In reaction to the motor theory and as a consequence of categorical results with nonspeech sounds and non-human listeners, an explanation was developed based on auditory sensitivities (e.g. Pastore, 1987; Stevens, 1981). In this view, categorical perception had nothing to do with any speech specific mechanisms, but was based on auditory threshold phenomena

(e.g. Pastore, 1987; Pastore et al., 1977; Schouten, 1980). According to Pastore, a stable and relatively precisely defined internal or external limitation gave rise to categorical perception. This limitation could be an internal psychoacoustic threshold or a constant reference stimulus.

However, there are several arguments against this auditory explanation. Firstly, in many nonspeech studies, the criterion for categorical perception is the phoneme boundary effect, which is not indicative of categorical perception according to the definition, since it ignores within-category performance. Classification and discrimination performance have to be strongly related both at the phoneme boundary and on both sides of it.

Secondly, it has been argued that categorical results with animals substantiate the general auditory explanation of categorical perception (e.g. Kluender & Lotto, 1994; Kuhl, 1981; Kuhl & Padden, 1982). Yet, the animal studies describe performance on only one of the two perception tasks: classification or discrimination. Therefore, these findings show only that animals (after prolonged training) can learn to divide a stimulus continuum between two speech sounds into two categories, but they tell us nothing about the correlation between classification and discrimination and thus about categorical perception.

Finally, it has been argued that the results of speech studies with non-humans do not necessarily imply that general auditory processes, or mechanisms common to humans and non-humans are at work. Non-human performance with speech is only analogous to human performance; it indicates that similar processes may arise from disparate evolutionary sources (Jusczyk & Bertoncini, 1988; Remez, 1989).

1.4 How robust are the categorical perception results?

As we have said before, numerous studies have claimed to have found evidence for 'categorical perception', but in reality these studies demonstrate different degrees of categorical perception. The differences between the empirical results have led to some discussion about the status of categorical perception (e.g. Cutting, 1982; Kluender, 1994; Massaro, 1987; Massaro & Cohen, 1983, Pisoni, 1973; Schouten, 1980, 1987).

1.4.1 The effects of phoneme class and stimulus naturalness

The most striking evidence for a large variation in the degree to which classification predicts discrimination is the one obtained as a function of phoneme class. The highest degrees of categorical perception have been demonstrated for the place and voicing contrasts between stop consonants (e.g. Abramson & Lisker, 1970; Liberman et al., 1961; Repp, 1981). Nasal consonants are perceived less categorically than stops, and liquids, semivowels, and vowels are perceived less categorically still (see review Repp, 1984). For fricatives the results are mixed: the fricative noises tested by Fujisaki and Kawashima (1970) were perceived highly categorically; however, Healy and Repp (1982) and Howell and Rosen (1984) found a low degree of categorical perception for their fricative noises. In short, this suggests that strongly categorical results are specific for stop consonants, with vowels at the other end of the scale. However, Pisoni (1973) found that relatively short vowels are perceived more categorically than longer vowels. Furthermore, Schouten and Van Hessen (1992) found a high degree of categorical perception for vowels, which they explain by the naturalness of the stimuli. This was confirmed by their study on the effect of stimulus quality on categorical perception: more natural stimuli were perceived more categorically than less natural stimuli (Van Hessen & Schouten, 1999). This suggests that stimulus factors may have an effect on the degree of categorical

perception. And more importantly, it indicates that the variation in categorical perception of different phoneme classes may diminish if, instead of synthetic, more natural stimuli are used. This motivated us to investigate the influence of stimulus naturalness on categorical perception. We compared categorical perception of vowel stimuli modelled on productions in isolated words with vowel stimuli that resembled productions in everyday speech. In particular, we compared categorical perception of stop consonants with the categoricalness of text vowels.

1.4.2 The effect of the discrimination task

Another source of variation is the discrimination task. One of the arguments against the categorical perception hypothesis has been the use of the ABX-discrimination task, which is the standard discrimination test for categorical perception research. This task is said to bias results towards categorical perception (e.g. Massaro & Cohen, 1983; Pisoni, 1975; Schouten, 1987). In the ABX task, listeners have to indicate whether stimulus X is identical to stimulus A or stimulus B. In doing so, they may try to remember both the auditory traces and the labels assigned to the A and B sounds. When X is presented, they try to match the sound of X with the auditory traces of A and B, but by this time they may have forgotten what A and B sounded like. If they have, the subjects must rely on the labels they have assigned to A and B and choose the one that matches the label they have assigned to X. This strategy is similar to classification and will be indicative of categorical perception. Therefore, the ABX task is not the most appropriate task to use when assessing categorical perception. However, the success of the ABX paradigm has led to its continued use. Only Crowder (1982), Pisoni (1973) and Repp, Healy and Crowder (1979) have compared ABX with a different discrimination task. Their results imply that categorical perception may depend, to some extent, on the type of discrimination task used. But, as Repp (1984) writes in his extensive review of the categorical perception literature, “A comparison of more than two paradigms for speech discrimination in a single study still remains to be done”. This is precisely what was done in the present study. Categorical perception was tested with five different discrimination paradigms, using the same vowel stimuli and the same subjects.

1.4.3 The effect of the inter-stimulus interval

The duration of the temporal interval (inter-stimulus interval, ISI) between stimuli to be discriminated has also been suggested to affect categorical perception results (Cowan & Morse, 1979; Pisoni, 1973; Van Hessen & Schouten, 1992). It has been proposed that, as a consequence of a relatively long ISI, there will be a decay of the auditory trace in short term memory, which will encourage listeners to base discrimination on phonetic representations in long term memory. This is one of the factors that contribute to the categorical results in the ABX task: ISI is often as long as 1 second. Studies that systematically varied ISI report that discrimination results show an increasing effect of phoneme labelling with a longer ISI (Cowan & Morse, 1979; Pisoni, 1973; Van Hessen & Schouten, 1992). In addition, the ISI effect appears to vary as a function of the discrimination task (Van Hessen & Schouten, 1992). It has to be noted that these studies only show the effect of ISI on listeners’ discrimination performance. Because no direct comparison is made between discrimination and classification, the effect on categorical perception remains speculative. This motivated us to conduct a series of experiments in which the influence of ISI on the relationship between

discrimination and classification was explicitly tested.

1.4.4 Other factors

Other factors that may influence the degree of categorical perception, but that have been reported in only few studies, are the individual listener (Pastore, Friedman & Buffato, 1976; Repp, 1981), instructions (Pastore, 1981), feedback (Hanson, 1977), and extent of training (Pisoni et al., 1982).

1.5 The present study

1.5.1 A test of the categorical perception hypothesis

The findings discussed above indicate that categorical perception results are not as robust as has often been claimed. Instead, it appears that the evidence for categorical perception in a particular experiment depends largely on specific choices made concerning the stimulus, the task, and task procedure. The variation in categorical perception led Cutting (1982) to conclude that "It makes it unworthy of being held in too high esteem as a touchstone for a specific process in speech perception". Before we concur with such a strong rejection, we want to investigate several factors that may contribute to this variation in a more systematic way than has been done before. Therefore, the causes of the variation and the consequences for the laboratory phenomenon that is called categorical perception form an essential part of the present study. Because there are so many factors influencing categorical perception, we have selected those that seem to have a relatively large effect: the stimulus, the task, and the inter-stimulus interval. The main research question and sub-questions that are addressed in the first part of this study are the following:

Is there an invariable relationship between discrimination and classification of a stimulus continuum between two phonemes?

- *Is there a difference in the degree of categorical perception of word vowels and text vowels?*
- *Are there differences in the degree of categorical perception of vowels as a function of the discrimination task?*
- *What is the effect of the duration of the inter-stimulus interval on discrimination and hence, on the degree of categorical perception of vowels?*
- *Is there a difference in the degree of categorical perception of text vowels and stop consonants?*

1.5.2 The development of phoneme categories

Another argument against a relationship between classification and discrimination of speech sounds may come from developmental studies. It has been shown that infants have the capacity to detect differences between syllables that differ in one phonetic segment, even between phonemes that do not occur in their native language. By 12 months of age, infants appear to have taken some steps towards the phonological categories of their native language: certain phonetic contrasts falling within the same native language category are no longer readily discriminated, despite the fact that the same contrasts were discriminated at 6 months of age (Jusczyk, 1992). This increased attention of the infants for native language contrasts does not preclude the discrimination of all non-native contrasts, as is also shown in cross-linguistic studies with adult listeners. The infant's acquisition of the native phonological system does not suggest that there is a loss of auditory sensitivity, but that there is a reorganisation of initial sensitivities (Jusczyk, 1992; Werker & Pegg, 1992). In recent infant research there has been a shift away from discrimination performance as a measure of language processing and language development. It is thought that, unless the infant is in a laboratory, discriminating phonologically minimal pairs is not the first, or most important, language task facing it; but learning words is. The fact that young infants can discriminate between /pE/ and /bE/ syllables does not necessarily reflect their ability to identify words that differ in this voicing feature (Swingley, 2000).

Unfortunately, there are only few studies that have investigated speech perception by young children of between 1 and 4 years of age (e.g. Barton, 1980; Svachkin, 1948/1973). At first sight, the results of these few studies seem to contradict the findings obtained with infants: the older children appear to be much worse at discriminating differences between syllables and words than the infants are. However, this contradiction is due to some important differences between the infant, child, and also adult studies. Firstly, different test procedures are used and secondly, the subjects' perceptual processes differ essentially from one another. In the infant studies, change-no change discrimination is used with words or syllables differing in one phoneme (e.g. ba-ba-ba-ba-da). The infants are trained to make a head turn when the stimulus changes. Discrimination with adult listeners normally involves a comparison of two or more stimuli presented within the same trial; the stimuli are taken from a continuum between two phonemes. Adult listeners are asked to indicate whether two stimuli are "identical" or "different" or to detect which stimulus differs from the other stimuli in a trial. Infants' early capacity to detect acoustic differences is assumed to be unrelated to representations of phoneme categories in their long-term memory, because they simply have not developed these representations yet. But in the case of adult listeners, the categorical perception hypothesis predicts that discrimination is mainly based on phoneme-labelling processes. For children of age 1 to 4, the story is different again. The *discrimination* response used with children usually involves a choice between one of two objects that have nonsense names or represent familiar words, differing only in one phoneme. This task is identical to two-alternative forced-choice *classification* in adult research: it does not measure to what extent children are actually able or unable to discriminate between speech sounds. This implies that these discrimination studies actually report children's ability to classify or recognise words, and this could very well be more difficult than detecting differences (the infant studies) between two acoustic signals that convey no meaning and have no lexicalised representations. In a recent Dutch study, the classification performance of 445 five-year-old children showed that there were several phonetic distinctions on which they made more than 20% errors (Stoep & Verhoeven, in press). In this study, a manual picture selection task was used with minimally differing word pairs that were familiar to the children (note that this

again is a classification task and not a discrimination task). The error percentages for /b - p/ and /d - t/ were as high as 43.1% and 32.7%. These results show that children of approximately 4 and 5 years of age have not fully acquired the classification system of their native language.

Other studies have shown that this is true even of children between 4 and 12 years of age. In these studies, the traditional classification task is used to assess categorical perception: a *single* acoustic cue is systematically varied along an acoustic continuum between two phonemes (e.g. Krause, 1982; Kuijpers, 1996; Mann, Sharlin & Dorman, 1985). The results demonstrate significant age-dependent differences in the phoneme boundary and the steepness/slope of the classification curve. There are reports showing that classification performance is adult-like early on (between 2 and 6 years) in development (Crul & Peters, 1978; Werker & Polka, 1993). Others claim that only between the ages of 10 and 12 years, phoneme boundaries and steepness of classification functions become adult-like (e.g. Burnham, Earnshaw & Clark, 1991; Kuijpers, 1996). It seems likely that this discrepancy between the reported acquisition ages reflects the fact that some phonemes are acquired earlier than others.

In the past fifteen years, there has been a growing interest in the development of the perceptual weighting of the *various* acoustic cues that specify a phonetic category (e.g. Morrongiello et al., 1984; Nittrouer, 1992, 1996, 2000). How do listeners classify speech sounds if there are no invariant cues that specify a certain phoneme category? It is assumed that listeners integrate various acoustic properties when they have to make a decision about a phoneme category. As part of the integration process, the adult listener has learned that some aspects in the acoustic signal are more important in signalling a certain phoneme than others are. Therefore, the listener assigns different weights to various acoustic properties. The acquisition of the appropriate weighting schemes for acoustic cues is essential for phoneme classification/categorical perception and thus for the comprehension of speech. But how are these weighting schemes acquired? In the present study, we try to shed some light on this aspect of speech and language development by addressing the following research question:

- *Is there a difference in the weighting of certain acoustic cues for the categorisation of speech sounds between adults and young children?*

1.6 Outline of the thesis

Chapter 2 is devoted to the difference in perception between word and text vowels. The effects of various discrimination tasks and of the duration of the inter-stimulus interval on categorical perception of vowels form the subject matter of Chapter 3. In Chapter 4, we compare vowel perception to stop consonant perception and test if the effect of the discrimination task on the degree of categorical perception differs as a function of these two types of phonemes. The results described in these three chapters are summarised and discussed in Chapter 5. The development of speech categories, especially children's weighting of certain acoustic cues for speech classification, is studied in chapter 6. Chapter 7 contains the general discussion and conclusion of the study described in this thesis.