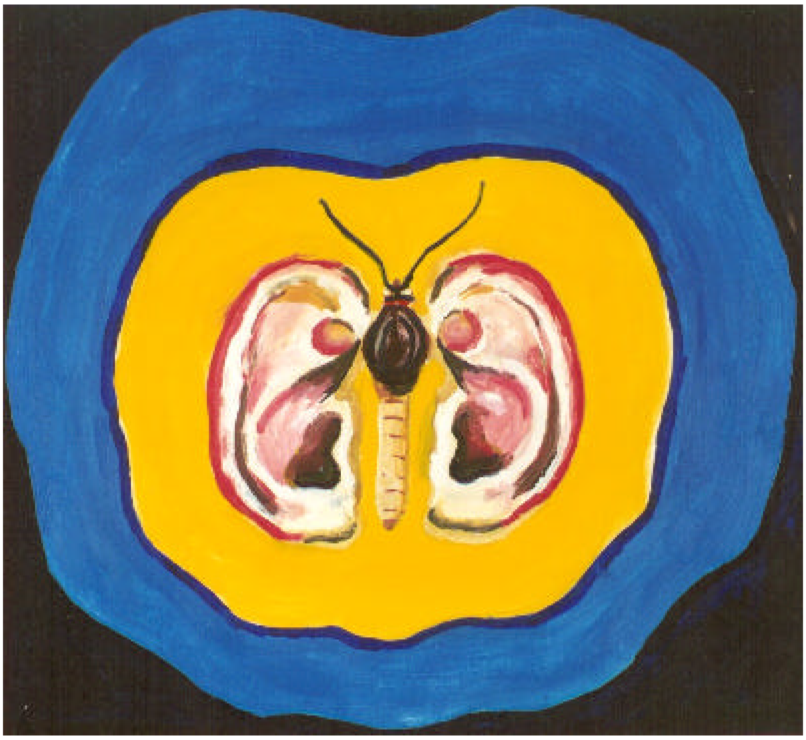
Ellen Gerrits

The categorisation of speech sounds by adults and children

A study of the categorical perception hypothesis and the developmental weighting of acoustic speech cues



LOT
Netherlands
Graduate
School of
Linguistics

Landelijke Onderzoekschool Taalwetenschap

The categorisation of speech sounds by adults and children

The categorisation of speech sounds by adults and children

De classificatie van spraakklanken door volwassen luisteraars en kinderen

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. Dr. H.O. Voorma
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 23 februari 2001
des middags te 16.15 uur

door
PETRA ARNO MARIA GERRITS
Geboren op 10 december 1967 te Helenaveen

Promotor: Prof. Dr. S.G. Nooteboom
Co-promotor: Dr. M.E.H. Schouten

Cover illustration: ‘*Als een vlinder* …’ by Jan Eduard.

ISBN 90-76912-08-4
NUGI 941

Printed in The Netherlands.

gedacht

Je hand is bijna je hond
je huid is bijna je huis
je vorm is bijna je worm
je gedicht is bijna wat je gedacht had

Gerrit Kouwenaar

Ellen Gerrits

The categorisation of speech sounds by adults and children

This thesis investigates the way adults and children perceive speech. With adult listeners, the question was whether speech is perceived categorically (**categorical speech perception**). With children, the question was whether there are age-related differences between the weights assigned to acoustic cues that specify certain speech contrasts (**cue-weighting**).

One goal of this thesis was to test the categorical perception hypothesis. According to this hypothesis, the acquisition of the native phonological system changes listeners' perception in such a way that they find it difficult to detect small differences between different realisations of the same phoneme, but relatively easy to detect equally small differences between realisations of two different phonemes. Several discrimination and classification experiments are described involving stimulus continua between vowels and between stop consonants. It is shown that categorical perception results are far from robust and that the degree of categorical perception is influenced by the discrimination task, the interstimulus interval, the listener, and the stimulus. Evidence is provided that listeners are perfectly capable of hearing small within-phoneme-category differences and thus that the acquisition of the phonological system does not have negative effects on the detection of differences between speech signals.

A second goal was to study the development of the perceptual integration of speech cues that specify a certain phoneme category. It was tested whether children of 4, 6, and 9 years old weight certain acoustic cues for stop consonants, fricatives, and vowels differently from adults. The results confirm earlier findings that children, especially the 4-year-olds, weigh certain speech cues more heavily, and other cues less heavily, than adults do. Children's ability to adjust the weighting of specific cues in the variable acoustic signal provides further evidence against the idea that acoustic detail in the signal becomes less detectable as language develops.

This study is of interest to phoneticians and psycholinguists, as well as to researchers working in the field of experimental speech and language development.

Contents

Acknowledgements xi

1 General introduction 1
1.1 Introduction 1
1.2 Categorical speech perception 2
1.3 Explaining the categorical perception results 3
1.4 How robust are the categorical perception results? 4
    1.4.1 The effects of phoneme class and stimulus naturalness 4
    1.4.2 The effect of the discrimination task 5
    1.4.3 The effect of the inter-stimulus interval 5
    1.4.4 Other factors 6
1.5 The present study 6
    1.5.1 A test of the categorical perception hypothesis 6
    1.5.2 The development of phoneme categories 7
1.6 Outline of the thesis 8

2 Categorical perception of word and text vowels 9
2.1 Introduction 9
2.2 Method 10
    2.2.1 Stimulus material 10
    2.2.2 Subjects 16
    2.2.3 Inter-stimulus interval 16
    2.2.4 The discrimination task 16
    2.2.5 Fixed and roving discrimination 18
    2.2.6 Design 18
    2.2.7 Procedure 19
2.3 Results 19
    2.3.1 Differences among subjects 22
    2.3.2 Control experiment 24
2.4 Discussion and conclusion 26

3 The effects of task and inter-stimulus interval on categorical perception 29
3.1 Introduction 29
3.2 Experiment I: 2IFC, 4IAX and 4I-oddity discrimination 30
    3.2.1 Method 31
        3.2.1.1 Stimulus material 31
        3.2.1.2 Subjects 31
        3.2.1.3 Design 31
        3.2.1.4 Procedure 31

3.2.2 Results 2IFC, 4IAX and 4I-oddity 32
3.3 Experiment II: Duration of the temporal interval between stimuli 36
3.3.1 Method 37
3.3.1.1 Design 37
3.3.2 Results long-ISI condition 37
3.4 Experiment III: AX and AXB discrimination 42
3.4.1 Method 42
3.4.1.1 Design 43
3.4.1.2 Procedure 43
3.4.2 Results AX and AXB 43
3.3 Discussion and conclusion 48

4 A comparison of categorical perception of stop consonants and vowels 51
4.1 Introduction 51
4.2 Method 53
4.2.1 Stimulus material 53
4.2.2 Subjects 54
4.2.3 Design 54
4.2.4 Procedure 54
4.3 Results 55
4.3.1 Stop-consonant results 55
4.3.1.1 Differences among subjects 59
4.3.1.2 Discussion stop-consonant results 60
4.3.2 Vowel results 62
4.3.2.1 Differences among subjects 63
4.3.2.2 Discussion vowel results 66
4.4 Discussion and conclusion 67

5 The categorical perception hypothesis revised 69
5.1 Introduction 69
5.1.1 Vowel naturalness 69
5.1.2 Discrimination task 70
5.1.3 Inter-stimulus interval 70
5.1.4 Stop consonants and vowels 70
5.2 A revised categorical perception hypothesis 71

6 The perceptual weighting of speech cues during childhood 73
6.1 Introduction 73
6.2 Perceptual integration of speech cues by children 74
6.2.1 Cue integration for stop consonants 74
6.2.2 Cue integration for fricatives 77
6.2.3 Cue integration for vowels 80
6.2.4 Summary of children’s perceptual integration of speech cues 80
6.3 Explaining children’s perceptual integration of speech cues 81
6.3.1 The development of speech perception 81
6.3.2 The development of the auditory system 82

        6.3.3 The development of the attention span 82
    6.4 The present study 83
    6.5 Method 85
        6.5.1 Stimulus material 85
            6.5.1.1 Fricatives 85
            6.5.1.2 Stop consonants 87
            6.5.1.3 Vowels 88
        6.5.2 Subjects 90
        6.5.3 Design 91
        6.5.4 Procedure 92
    6.6 Results 93
        6.6.1 Statistical analysis 93
        6.6.2 Fricative results 94
        6.6.3 Stop-consonant results 97
        6.6.4 Vowel results 99
    6.7 Discussion 102
        6.7.1 Cue weighting for consonants 102
        6.7.2 Cue weighting for vowels 103
        6.7.3 The development of the lexicon 104
        6.7.4 Further research 105
            6.7.4.1 The development of phonemic awareness 105
            6.7.4.2 Language disorders 106
    6.8 Conclusion 107

7 General discussion and conclusion 109

References 113

Appendix A Pictures of stimulus words **123**
Appendix B Tables with results /ΣΟκ/-/σΟκ/ **124**
Appendix C Tables with results /πΟπ/-/κΟπ/ **125**
Appendic D Tables with results /ζα κ/-/ζΑκ/ **126**

Samenvatting (Summary in Dutch) **127**

Curriculum Vitae **133**

Acknowledgements

Since some of you will only read this part of my thesis, I start with a brief summary of the results presented in detail in the following chapters. The aim of the present study was to reveal some aspects of speech perception, especially …. okay …, I'll end here.

In this section, I want to acknowledge a few of all the people that have helped and supported me while I was working on this research project.

First of all, I sincerely thank my supervisors at the Utrecht Institute of Linguistics OTS, Sieb Nooteboom and Bert Schouten. Sieb in particular for his comments and suggestions in the past years and for his constructive criticism on earlier versions of this thesis; Bert for cheering me up when needed and for his patience and persistence in supervising on a regular basis (which included serving me afternoon tea). I want to thank the members of the supervising committee (Aad Houtsma, Cecile Kuijpers, and Frank Wijnen) for their help during various stages of the project. In addition, I am grateful to the members of my thesis committee (Vincent van Heuven, Aad Houtsma, John Kingston, Louis Pols and Frank Wijnen) for reading the manuscript and for their useful comments.

In the ideal linguistic universe, all university students and employees would line up and wait their turns to provide the investigator with fascinating data. Anyway, among the few who will attain immortality as the participants in my lasting experiments are my (ex)-colleagues Sergio Baauw, Esther Janse, Olaf Koeneman, Sandra Peters, and Maarten Janssen (who turned out to be an auditory genius). Furthermore, I thank the children and the staff of elementary school "De Ludgerschool" and of day-care centre "De Vlieger", for their co-operation in the experiments.

To run experiments, it is necessary to create an on-line listening task and for this part I thank Theo Veenker. He provided technical support when needed and was of great help, especially when one of the older terminals sort of 'exploded' during testing.

My colleagues at the Utrecht Institute for Linguistics OTS are thanked for their interesting discussions and small talk during numerous coffee breaks and lunches. Special thanks are for my (ex)-colleagues: Sergio Baauw, Elma Blom, Gerrit Bloothooft, Christien Bok, Remca Burger, Jan Don, Paz Gonzalez, Esther Janse, Maarten Janssen, Jan de Jong, Lisanne Klein Gunnewiek, Olaf Koeneman, Esther Kraak, Guus de Krom, Niels van der Mast, Anna Mlynarczyk, Iris Mulders, Brigit van der Pas, Wim Peeters, Sandra Peters, Mirna Pit, Hugo Quené, Saskia te Riele, Maya van Rossum, Nada Vasic, Willemijn Vermaat, Frank Wijnen, Carien Wilsenach and Inge Zwitserlood. In addition, a lot of thanks are for Saskia & Sandor, David van Kuijk, Lian Nijland, Robert Eklund, and Paola Escudero, who were great 24-hours company during several conferences abroad.

A warm thank you goes out to my brother Arno and sister Hanneke, Jan and José, my family, and all my personal friends for their interest and listening ears, their good company, and for asking whether I would ever stop with this student thing and get a real job. I end with thanking my parents for their unfailing belief in me and for an endless number of supporting phone calls; this book is dedicated to them.

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 /τΣΑ-ΣΑ/, 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.

Chapter 2

Categorical perception of word and text vowels[1]

2.1 Introduction

Several studies have shown that there is a difference in the degree of categorical perception between stop consonants and vowels. Stop consonants are said to be more categorically perceived than vowels (e.g. Fry et al., 1962; Pisoni, 1973; Repp, 1981; Stevens et al., 1969).

This difference in perception has received various explanations. Early explanations, based on the Motor Theory of Speech Perception, posit that it is the result of a difference in the underlying articulatory gestures (Liberman et al., 1967). Perception is categorical when the articulatory space is relatively discontinuous. Stop consonants show marked acoustic variability, but speakers seem to use discrete places of production for them. Vowel articulation is much less discrete: there is a fair amount of articulatory overlap between adjacent vowel categories (Liberman et al., 1967). However, the mechanisms by which perceptual processes might refer to articulation have always remained obscure, and this has led many researchers to dismiss the motor theory (see review Repp, 1984).

A second explanation for the processing differences between vowels and stop consonants is given by Ades (1977), Rosen and Howell (1987) and Sachs (1969). They speculate that the differences arise from differences in the perceptual distances between stop consonants and between vowels: two adjacent vowels span a wider perceptual range in terms of JND's than two adjacent consonants. Thus there should be a larger range of ambiguity along a vowel continuum than along a consonant continuum. One would therefore expect a smaller discrepancy between classification and discrimination (in other words, the results

---

[1] This chapter is an adapted version of Gerrits and Schouten (1998).

would be more categorical) for consonants than for vowels. However, the results of Ades (1977) and Macmillan, Goldberg and Braida (1988) revealed no difference in perceptual range in their stimulus sets: the total number of JND's between /ba/ and /pa/ was about the same as between /i/ and /I/.

A third explanation is proposed by Pisoni (1973, 1975), Sachs (1969), and Tartter (1981, 1982). They explain the difference in perception between stop consonants and vowels by the difference in cue duration. The essential acoustic cues for stop consonants are rapidly changing F1 and F2 transitions and a brief noise burst (Liberman et al., 1954; Tartter, 1981, 1982). By contrast, vowels are assumed to remain uniform over a much longer duration (Delattre et al., 1952). This difference in duration between vowels and stop consonants has an effect on the availability of auditory memory for these two classes of speech sounds. According to Pisoni (1973), the outcome of a stimulus comparison during discrimination depends on two types of memory components: trace memory and (phonetic) label memory. This is in line with the dual-process model of Fujisaki and Kawashima (1971), who propose that discrimination may be performed in a trace comparison or in a labelling mode. They explain the difference in perception between stop consonants and vowels by differences in the duration of the critical information in these speech signals. The short cue duration in consonants is responsible for the inferior performance of trace memory. Presumably, the decay of rapidly changing acoustic information is too fast to make an acoustic comparison of consonantal stimuli possible, with the result that discrimination is performed in a phonetic mode. This is not the case for vowels produced in isolated words (citation form).

In line with this explanation, Schouten and Van Hessen (1992) propose that the lack of categorical perception of vowels may be due to the nature of the stimulus material that has been used. Up to now, the vowels used as stimulus material have been modeled on productions in isolated words. When produced in isolated words, vowels will be lengthened (overarticulation). The duration of a plosive will be less affected by overarticulation than that of a vowel: the articulation of the burst can hardly be lengthened and the length of the formant transitions is also restricted since too much lengthening would lead to a change in phoneme identity. It is hypothesised that in running speech temporal reduction and more complex spectral coding of the vowel will make vowel perception more categorical (see also Pisoni, 1973; Repp, 1984; Sawusch, Nusbaum & Schwab, 1980; Schouten & Van Hessen, 1992; Stevens, 1968; Studdert-Kennedy, 1976; Tartter, 1981). The shorter duration and the more complex coding of vowels is expected to force listeners to make a quick decision about the phoneme category, especially when stimuli are difficult to discriminate. This will strengthen the relationship between discrimination and classification of the same stimuli. To test this hypothesis, the difference in perception was studied between vowels spoken in isolated words and vowels in a text read at a fast rate. The question of interest was whether text vowels would be perceived more categorically than word vowels.

2.2 Method

2.2.1 Stimulus material

The first step in stimulus generation was recording the vowels /u/ and /i/ in the meaningful words /pup/ and /pip/ produced both in isolation and in a text that was read aloud at a rapid speech rate by a male native speaker of Dutch. The vowels produced in the isolated words will be called 'word vowels' and the vowels produced in the words in the text will be referred to as 'text vowels'. Since the perception of naturally produced vowels was of interest in this study, one could argue that the ideal stimulus material should be based on vowels from spontaneous speech. It was decided to elicit vowels in a rapidly read text in order to control the stimulus material, on the assumption that fast reading produces the same type of underspecification as spontaneous speech. The vowels /u/ and /i/ were selected, because it was expected that the differences between the two speech conditions would be greater with these two vowels than with most other vowels, due to the relatively long articulatory trajectories required to reach them.

There were nine repetitions of /pup/ and /pip/ in the list of words and in the text. Five phonetically untrained listeners identified 30 ms segments of these vowels in an open-set identification task. The vowels spoken in isolation were significantly more often identified as /u/ and /i/ than the vowels from the fast read text (65% versus 32% for /u/ and 42% versus 7% for /i/). Frequently used other response categories were /o/ for /u/ and /y/ for /i/. The word pairs that were used as endpoints in the two stimulus continua were selected by a listening panel that consisted of five phoneticians. They rated the word pairs on a 7-point acceptability scale. The most acceptable ones were selected for the experiment.

Description of the original vowels

To see if there were any differences between the vowels spoken in isolated words and fast text, the durations and formant frequencies were measured. The text vowels were temporally reduced compared to the word vowels. Vowel duration was defined as the duration of the voiced portion between the two /p/'s. The steady-state part of a vowel was defined as that part during which the frequency of F2 did not differ by more than 50 Hz in successive analysis frames, which were 16 ms apart and 32 ms wide. The rest of the vowel consisted of formant transitions.

The duration of the word vowel /u/ was 90 ms, the steady-state component was 60 ms. The duration of the text vowel /u/ was 70 ms, with a steady-state component of 30 ms, a reduction of 22% and 50%, respectively. The duration of the word vowel /i/ was 90 ms, the steady-state component was 50 ms. And the duration of the text vowel /i/ was 70 ms, with a steady-state component of 15 ms, a reduction of, respectively, 22% and 70%. This temporal reduction is comparable to the reduction reported in studies by Schouten and Pols (1979) and Van Son and Pols (1990). The temporal vowel reduction found by Schouten and Pols (1979) was 28%. The reduction of steady-state segment duration was on average 38%. Van Son and Pols (1990) found a temporal reduction of 15% between vowels in a text that was read at a normal speech rate and as fast as possible.

In addition to the duration measurements, an analysis of the formant frequencies of the vowels in the two speech conditions was performed. Since automatic formant extraction failed with the text vowels, formant frequencies in all vowels were estimated by eye. Figure 2.1 shows the spectral envelopes of the word vowels /u/ and /i/, figure 2.2 does the same for text vowels /u/ and /i/. The FFT analysing window was 25 ms and contained the highest amplitude in the vowel. The spectrum was cepstrally smoothed with a cepstral lifter of 5 ms in the quefrency domain.

For the word vowels there were no difficulties in locating the formants in the frequency spectrum. The curves in figure 2.1 show resonance peaks that can be interpreted as formants. The estimated values of the first three formants are 370 Hz, 683 Hz, and 2086 Hz for /u/ and 284 Hz, 2065 Hz, and 2741 Hz for /i/. This is in accordance with the formant frequencies of Dutch /u/ and /i/ measured by Pols, Tromp, and Plomp (1973) and Van Son and Pols (1990).

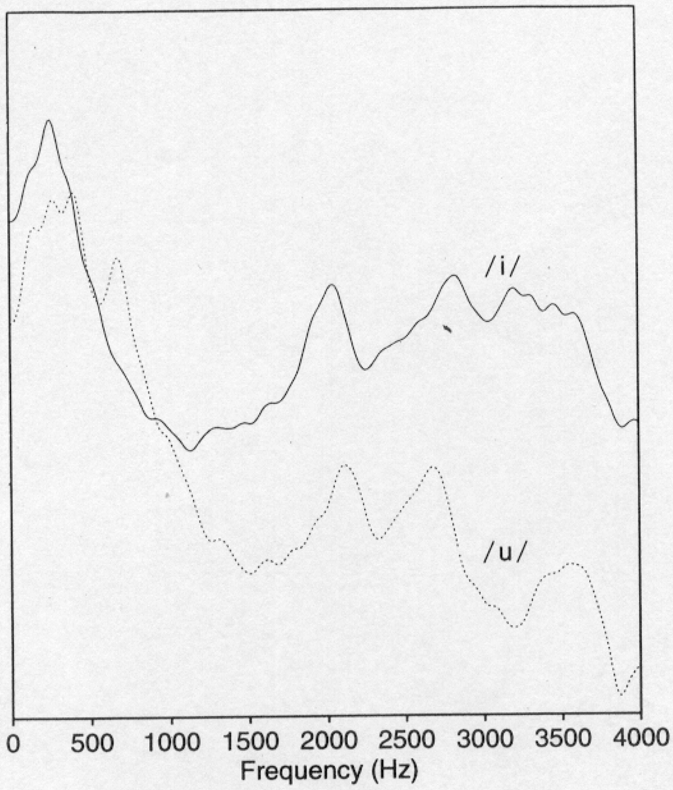


Figure 2.1 Spectral envelopes for the word vowels /u/ and /i/.

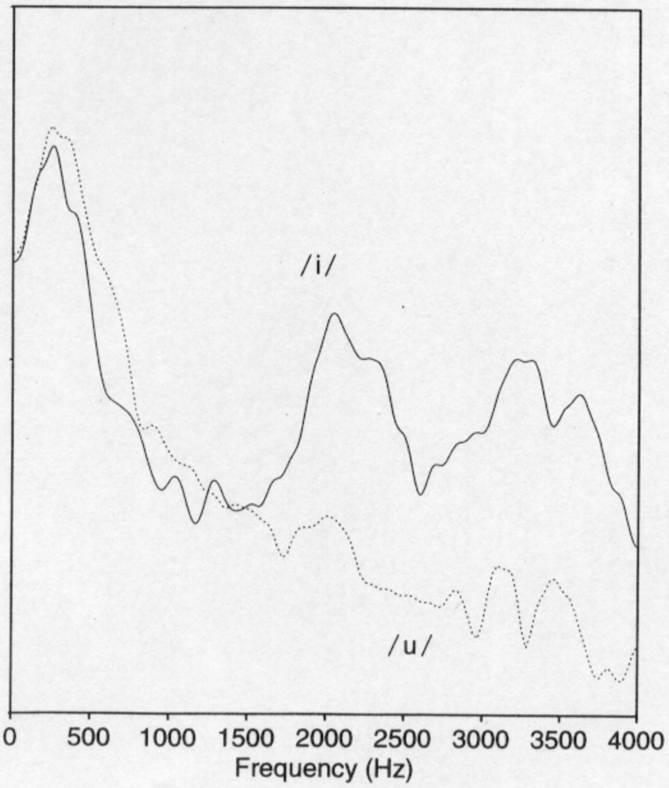


Figure 2.2 Spectral envelopes for the text vowels /u/ and /i/.

In figure 2.2, the spectral envelopes of *both* vowels show amplitude peaks around 250 Hz, 2100 Hz, and 3200 Hz, although at different relative amplitudes. If there had been no a priori knowledge of where the formants should be located, these peaks might have been interpreted as representing the first three formants. Despite the spectral similarity between the two curves in figure 2.2, the vowel that is represented by the dashed line clearly sounds like an /u/, and the vowel of the solid line is without doubt an /i/. Ladefoged (1967) noted that it is often impossible to locate the second formant when it is close to the first formant and lower in intensity. Either the lowest harmonics constitute two formants, or there is no second formant with sufficient intensity for its frequency location to be measured. Ladefoged (1967) added that when F2 is low in intensity, F3 is often impossible to locate for the same reason. Both Ladefoged (1967) and Pols et al. (1973) write that without knowing what vowel phoneme the speaker intended, and without previous knowledge of where the formants of that vowel should be, it is impossible to make valid measurements of the formant frequencies. The spectral data of our text vowels seem to be a nice illustration of their observations.

In short, the duration and formant frequency measurements show that there is a temporal reduction of the text vowels and that there are spectral differences, notably a clustering of formants for the text vowels.

Synthesis of the vowel continua

The interpolation method used had been developed for the study of categorical perception by Van Hessen (1992). Stimuli in a continuum between two natural utterances are obtained by interpolation between the relative amplitudes of the spectral envelopes of these utterances. Parameters such as fundamental frequency, duration, and voice quality, remain constant. The first step in this interpolation method is a separation between the source (vocal folds) and the filter (vocal tract) of the original speech signals. Only the filter characteristics are used for interpolation. The second step is an analysis of the spectral envelopes of the original words in terms of the phases and amplitudes of a large number of spectral components between 80 and 8000 Hz, depending on spectral density. By means of linear interpolation of the amplitudes and interpolation of the phases, spectral envelopes are constructed for the stimuli in between the original words. In the last step, each new spectral envelope (filter) is multiplied with the spectrum of the source of one of the original speech signals. The reconstructed endpoint signals are nearly indistinguishable from the original ones. More details of this procedure are described in Schouten and Van Hessen (1992), or Van Hessen (1992). For the interpolation between the vowels in the present study, the spectral envelopes were analysed in phases and amplitudes of 70 spectral components. The spectral envelopes of the eight stimuli, obtained by means of seven linear interpolation steps between each of the 70 pairs of spectral components, were then reconvolved with the original source spectrum of the /u/. The interpolation was always done in overlapping 25.6-ms time frames over the full length of the vowel (frame shift was 6.4 ms).

A disadvantage of the interpolation method is that it generates a speech continuum which has no correlate in speech production. Usually, a speech stimulus continuum is generated with an acoustic parameter that does have articulatory meaning, such as a formant frequency. Nevertheless, for our vowel study the amplitude-interpolation method was preferred to working in the formant domain. In this way the risk was avoided that, after having listened for a while to stimuli in which only one or two parameters were varied, some subjects would learn to attend selectively to those parameters. The experimental design was intended to motivate the listeners to focus on the speech signal as a whole. (Van Hessen and Schouten, 1999, have shown that there is an increase in categorical perception as synthesis quality improves from a simple synthesis by rule, via LPC synthesis, to the more complex method used in the present study.) The importance of stimuli in which more than one parameter is varied is also mentioned by Repp (1984) and Liberman (1996), who predict that, with proper synthesis, when the acoustic signal changes in all relevant aspects and not just one cue is varied, the discrimination functions will come much closer to being perfectly categorical. Moreover, since no second formant could be defined, interpolating the formants of the text vowels was impossible.

Stimulus continua

Stimulus generation resulted in two continua of 8 stimuli that sounded completely natural and convincingly like utterances from the original speaker. The continua were named as follows: 1) Word Vowel Continuum; 2) Text Vowel Continuum. In each continuum, the initial and final /p/ of the stimuli were copied from the original word /pup/. In a pilot experiment the stimuli of the two continua were identified (open set) by a listening panel that consisted of five phoneticians, all well-trained listeners. The listeners' identification responses were always /u/ or /i/. In none of the cases were the stimuli identified as the Dutch vowel /y/, which might have been possible because F2 of this central vowel is in between F2 of /u/ and /i/. The absence of an intermediate /y/ is an effect of the interpolation method.

Fundamental frequency and duration of the stimuli were the same as those of the original /pup/. In the word vowel continuum, average *F*0 was 120Hz and stimulus duration was 215 ms (vowel 90 ms, steady state 60 ms). The duration of the stimuli in the text vowel continuum was 187 ms (vowel 70 ms, steady state 30 ms) and *F*0 was on average 125 Hz.

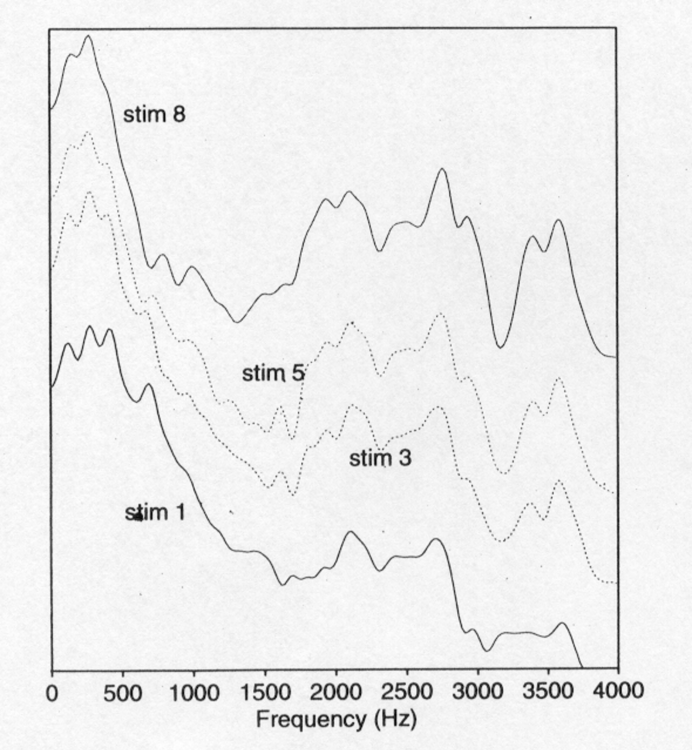


Figure 2.3 Spectral envelopes for stimuli 1, 3, 5, and 8 of the word vowel continuum.

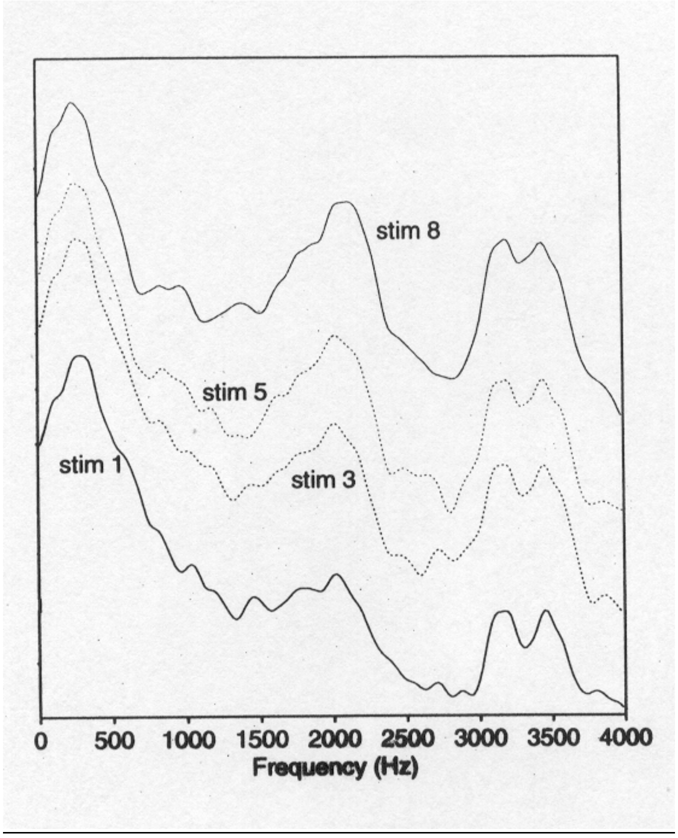


Figure 2.4 Spectral envelopes for stimuli 1, 3, 5, and 8 of the text vowel continuum.

The spectral differences between the two stimulus continua are illustrated in figures 2.3 and 2.4. These figures show the spectral envelopes of stimuli 1, 3, 5, and 8 of both the word and the text vowel continuum. The various curves have been shifted along the vertical axis for maximum clarity. In both figures stimulus 1 sounded like /u/ and stimulus 8 sounded like /i/. The FFT analysing window was 25 ms and contained the highest amplitude in the vowel. The spectrum was cepstrally smoothed with a cepstral lifter of 5 ms in the quefrency domain.

In figure 2.3 it can be seen that the spectral envelopes of stimuli 1 and 8 (the endpoints) are identical to those of the original vowels in figure 2.1. It is clear that the differences between the spectra of the endpoint stimuli and the intermediate stimuli (only 3 and 5 are plotted) consist of shifts in the relative amplitudes of the spectral envelope of the vowel in stimulus 1 towards the relative amplitudes of the spectral envelope of the vowel in stimulus 8.

In figure 2.4 it is shown that the spectral envelopes of the endpoint stimuli have the same shape as the envelopes of the original vowel in figure 2.2. The spectral envelope of stimulus 1 has no significant amplitude peaks that indicate the presence of a second formant for /u/, and in the spectrum of the vowel in stimulus 8 there is no amplitude peak that can be unambiguously classified as the third formant for /i/. Nevertheless, stimulus 1 did sound like /u/ and stimulus 8 sounded like /i/. Figure 2.4 also shows that the stimuli between the endpoints differed in the relative amplitudes of the spectral envelopes.

2.2.2 Subjects

The subjects were 19 students of the Faculty of Arts at Utrecht University. They had no known hearing deficits and were all native speakers of Dutch. They were paid a fixed hourly rate.

2.2.3 Inter-stimulus interval (ISI)

Since it was desirable to encourage a comparison of the acoustic cues in stimuli during discrimination and since the auditory trace of speech sounds is time-dependent, it was important to make a considered decision about the inter-stimulus interval. If the inter-stimulus interval exceeds the duration of the auditory trace of the vowel stimuli, all that is left of the first stimulus is a representation coding the relationship of the stimulus to the other stimuli in the experiment, or to pre-established categories, or to both (Pisoni, 1973). Therefore, an interval was preferred that did not exceed the auditory trace of the stimulus. Massaro (1972a, 1972b, 1974) tried to determine the time a sound pattern is held in some preperceptual form. His results indicated that the auditory trace of an acoustic signal has faded after approximately 250 ms, which is in agreement with the findings by Dorman et al. (1977) and Plomp (1964).

Cowan and Morse (1986), Pisoni (1973), and Van Hessen and Schouten (1992) tested the effect of a varying inter-stimulus interval on discrimination performance. On the basis of Massaro’s results, *within-category* discrimination should decrease with increasing interval, reflecting the fading of the trace. This is in agreement with the results of Van Hessen and Schouten (1992) who found a decrease of within-category stop-consonant discrimination with an increase in ISI from 100 to 300 ms. The *across-category* results of the discrimination studies by Cowan and Morse (1986), Pisoni (1973), and Van Hessen and Schouten (1992) confirmed the notion that processing of the auditory signal is not terminated after 100 to 200 ms. Their results indicated that if listeners use a labeling strategy to compare the stimuli, discrimination improves as an effect of increasing ISI: discrimination performance increases rapidly between 100 and 500 ms, reaches a maximum between 500 and 1000 ms and falls gradually as ISI increases further.

On the basis of these results it was assumed that, after an ISI of more than 100 - 200 ms, labeling processes would take over from trace coding. An interval shorter than 200 ms was not used, since this might increase the chance of mutual masking among the stimuli. In line with Massaro (1972a, 1972b, 1974), Pisoni (1975) and the within-category results of Van Hessen and Schouten (1992), it was therefore decided to use 200 ms intervals to make sure that trace information was available for direct auditory comparison of successive stimuli in a trial.

2.2.4 The discrimination task

The prediction test is the most widely used formal criterion of categorical perception (the Haskins model). It requires a close correspondence between the actual discrimination of speech stimuli along a continuum and discrimination performance predicted from classification results. The procedure of the prediction test has been criticised by Massaro and Cohen (1983). The prototypical discrimination test used to assess categorical perception is the ABX task (Liberman et al., 1957). In view of the relatively short time span of trace memory, however, the observed phenomenon of categorical perception may, according to Massaro and

Cohen (1983), reflect the exclusive use of phonetic memory. Subjects in the ABX task 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 will produce the results usually attributed to categorical perception. Pisoni (1975) and Schouten (1987) also note that the ABX task prevents a direct comparison between successive stimuli and forces the listener to use an encoded categorisation in discrimination. Therefore, the ABX task is not the most appropriate task to use when assessing categorical perception.

The same problems as for ABX discrimination may hold for another paradigm: 2IFC discrimination. In this paradigm, two different stimuli are presented, AB or BA. The subject has to determine the order in which the stimuli are presented. This task has a long tradition in research on intensity resolution (e.g. Macmillan et al., 1977). In intensity experiments, the listener can refer to the order as "soft-loud" or "loud-soft". In the case of speech stimuli, it is necessary to explain to the subjects what the term "order" means and to mention the phoneme categories in the instructions, e.g. /u-i/ and /i-u/. This is at the risk of encouraging labeling behaviour.

To avoid strategies that rely exclusively on labeling, it was necessary to opt for a task that reduced the load on auditory memory and encouraged a direct auditory comparison between the stimuli in a trial (Massaro & Cohen, 1983). Such a task is AX (Same-Different) discrimination, in which AA, AB, BB, and BA combinations are presented. A disadvantage of this paradigm is that, when subjects are asked if two stimuli are different or equal, they may decide to respond "different" only if they are very sure of their decision. This means that AX is not bias free: a subject's response is determined by a subjective criterion (i.e. a boundary between two categories) on a scale between "same" and "different".

A discrimination test that has been shown to be more sensitive to acoustic differences between speech stimuli is the 4IAX task (Pisoni, 1975). The trials in this task consist of eight combinations: ABAA, BAAA, AAAB, AABA, and BABB, ABBB, BBBA, BBAB. The temporal interval between stimuli 2 and 3 is longer than between the other stimuli, so that listeners hear two pairs within one trial. They have to decide which pair contained identical stimuli, pair 1 or pair 2. It is assumed that listeners first determine the differences between the stimuli within the pairs and, in a second step, determine which of the two differences is the smaller one. The correct decision is based on trace information and does not involve subjective criteria.

Yet, it was decided not to use a 4IAX task, but 4I-oddity. In 4I-oddity, the A and B stimuli are presented randomly in the two orders AABA or ABAA, with a 50% a priori probability. Stimulus A at the beginning and end of each quadruplet functions as a reference. Listeners have to respond by indicating if the 'oddball' (stimulus B) is in the second or third stimulus interval. In principle this task is as bias free as 4IAX, and it has a much shorter experimental duration. However, although it is a four-interval task, the optimal decision rules defined by Macmillan and Creelman (1991) predict that the ideal observer will ignore the reference stimuli (interval 1 and interval 4) and perform the 4I-oddity task as if it is a standard 2IFC task (see also Heller & Trahiotis, 1995; Trahiotis & Bernstein, 1990; and Verbunt, 1996). The advantage over 2IFC is that listeners *can* decide about the oddball without needing to refer to any internal criterion. In short, it is expected that 4I-oddity combines some of the important aspects of 2IFC and 4IAX and that listeners will have the choice between two perceptual strategies: a 2IFC-like phoneme-labelling strategy and a 4IAX-like trace-coding strategy. This is also the conclusion drawn by Heller and Trahiotis (1995): their 4I-oddity results indicate that listeners do not always use the optimal listening strategy but also use the reference stimuli to reach a decision about the oddball. In line with these results it is predicted that the use of the reference stimuli is optional and therefore that 4I-oddity will be, hopefully, neutral with respect to auditory processing and phoneme labelling. This neutral position should make 4I-oddity a proper diagnostic instrument for determining whether categorical perception occurs.

2.2.5 Fixed and roving discrimination

The stimuli were presented in a 'fixed discrimination' and a 'roving discrimination' test (Macmillan, 1988). In the fixed context, listeners have to respond to the same stimulus pair repeatedly during a block of trials before the next pair (which is chosen at random) is presented in another block. This means that within one block the stimulus range is very small, which reduces stimulus uncertainty. Listeners know which stimulus pair will be presented within the block and are consequently able to build up a temporary representation of the two stimuli (Durlach & Braida, 1961). Discrimination performance in the fixed context is expected to be higher than in the roving context.

In the roving context stimuli were drawn randomly from the total stimulus continuum. For instance, if there are 8 stimuli, the first trial may use 4 and 5, the second 1 and 2, etc. The roving context is the one almost invariably used in speech discrimination experiments. The stimulus range in this context is the whole continuum and therefore stimulus uncertainty is much higher than in the fixed context. As a consequence of high stimulus uncertainty it is expected that listeners are unable to form temporary labels for each of the speech stimuli and therefore that labelling effects in the roving data reflect the use of category labels stored in long-term memory.

2.2.6 Design

The experiment consisted of six tests, three for each of two vowel continua, involving the same subjects. The tests were fixed and roving 4I-oddity discrimination and classification. The subjects took the tests in a fixed order: the classification experiment was always performed after the discrimination tests.

The discrimination experiment used was the 4I-oddity task (AABA/ABAA). In this test, the subjects' task was to indicate whether stimulus B (the oddball) occurred in the second or the third interval. The stimuli in the second and third intervals always differed by one step along the continuum; the number of comparisons was therefore seven. The inter-trial interval was determined by the response time. The inter-stimulus interval within a trial was 200 ms.

The fixed test consisted of 7 blocks, one for each comparison, which were clearly separated from each other. Each block contained 64 trials, 32 for each of the two possible combinations, AABA and ABAA. In the roving discrimination test, 7 x 64 trials were presented randomly.

In the classification test, each stimulus had to be identified 64 times in a random order. Classification involved a forced choice between two alternatives, the vowels /u/ and /i/. There was no response time limit.

2.2.7 Procedure

The stimuli were presented to the subjects over headphones in a sound-treated booth. In the discrimination tests, it was stressed that differences between the stimuli would be small, and in most cases could only be detected by listening carefully to all details of a stimulus. The subjects responded by mouse-clicking on one of two response fields (labelled “2” and “3”) on a computer screen. After the response had been made, visual feedback (250 ms) of the correct answer was given so that the subject was able to judge and possibly improve his or her performance. Discrimination training consisted of 126 trials, and was intended to familiarise subjects with their task. In the fixed discrimination context the first ten trials of every block were considered practice and were not included in the data analysis.

In classification, one stimulus was played on each trial, and the subject had to identify it by mouse-clicking on a response field labelled “oe” or “ie” (/u/ and /i/). The only training consisted of 16 trials.

2.3 Results

The discrimination and classification results were computed in terms of d' (e.g. Kaplan, Macmillan & Creelman, 1978; Macmillan & Creelman, 1991). First, the proportions of hits (H) and false alarms (FA) of the adjacent stimulus pairs on the continuum were transformed in z-scores and, second, the z-scores were converted into d' according to the decision model appropriate for the task. The 4I-oddity discrimination d' scores were calculated by subtracting z(FA) from z(H) with a standard deviation $\sigma = \sqrt{2}$ (Macmillan & Creelman, p. 121, 1991). The results of the classification test are presented as predicted discrimination scores. The transformation of the classification data into predicted discrimination was done as follows. For each pair of stimuli AB and response alternatives /u/ and /i/, the proportion of /u/-responses to stimulus A (position n) was regarded as an estimate of p(H) and the proportion of /u/-responses to stimulus B (position n+1) was taken as an estimate of p(FA). The classification d'-values were determined by subtracting z(FA) from z(H). The values of p(H) and p(FA) were forcibly limited to the 0.99 to 0.01 range, which meant that the maximum d'-value that could be obtained was $4.65 * 0.5\sqrt{2} = 3.29$ for oddity discrimination and 4.65 for classification.

The results are displayed in figures 2.5 to 2.7. Figures 2.5 and 2.6 show the classification and discrimination data for the word vowels and the text vowels, respectively. In figure 2.7, the results are recalculated to get a better indication of the difference between the two vowel conditions. The data in the figures represent the averages of 19 subjects’ individual d'-scores. The numbers (n) along the abscissa refer to stimulus pairs, consisting of stimuli (n) and (n+1); n is therefore a number between 1 and 7. The d' result at stimulus pair 6, for example, represents the discrimination of stimulus 6 and stimulus 7. Stimuli in pair 1 resemble /u/ and stimuli in pair 7 sound like /i/.

In figure 2.7, the two vowel conditions (displayed in figures 2.5 and 2.6) are compared by calculating the difference scores between obtained discrimination and discrimination predicted by the classification data for each stimulus pair (as in Pisoni, 1975). If text vowels are perceived more categorically than word vowels, a smaller difference is expected between the obtained and predicted functions for the former condition than for the latter.

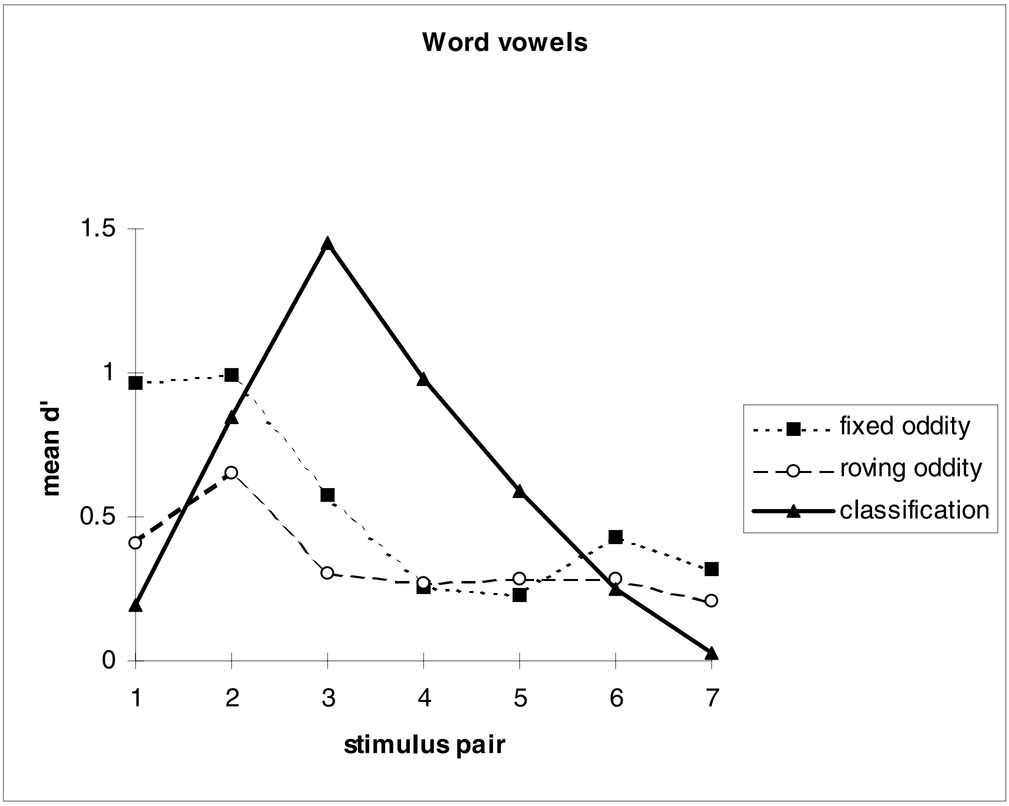


Figure 2.5 Classification and discrimination results for the word vowels.

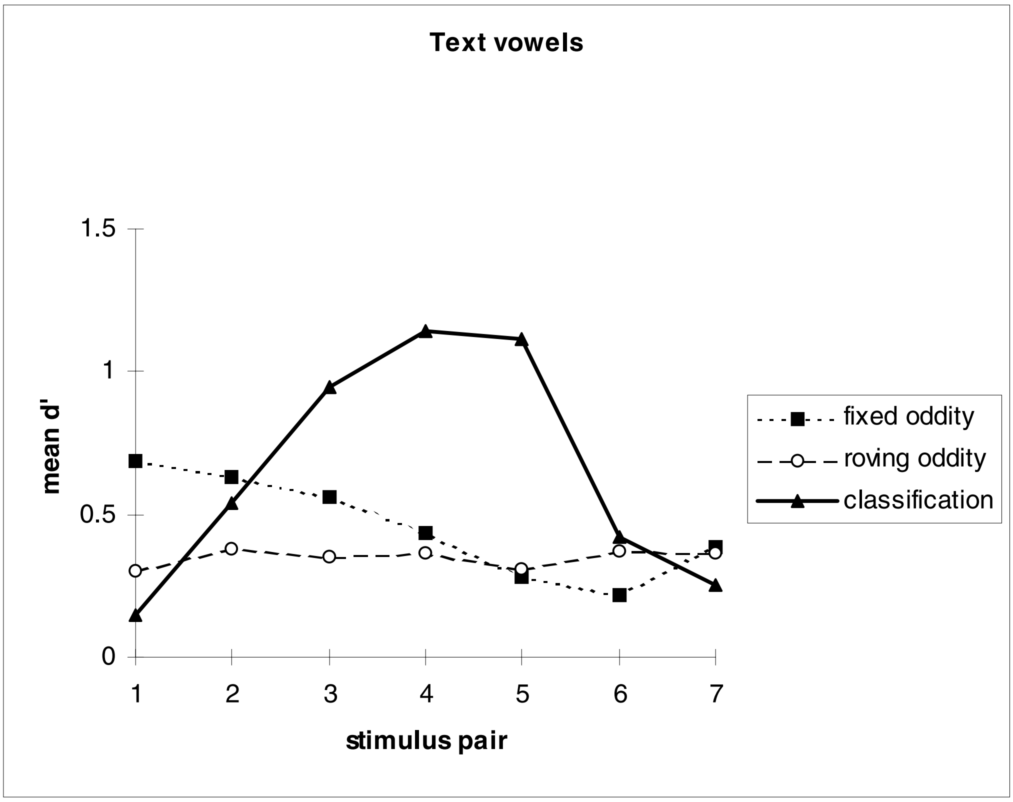


Figure 2.6 Classification and discrimination results for the text vowels.

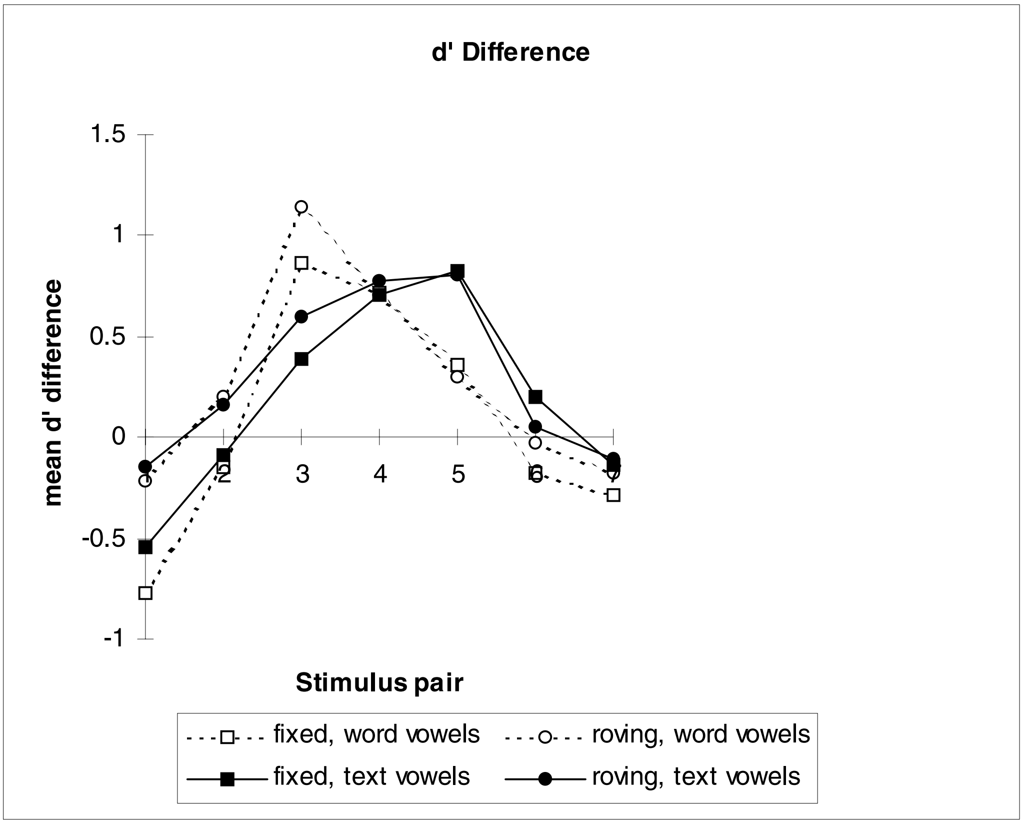


Figure 2.7 Difference scores between the obtained and predicted data for each stimulus pair in figures 2.5 and 2.6. The dashed lines represent the difference scores for the word vowels, the solid lines represent the difference scores for the text vowels.

A series of paired samples sign tests on the d'-difference scores between classification and discrimination in figure 2.7 revealed that there were only two stimulus pairs with a significant effect of vowel condition: stimulus pairs 3 and 5. At stimulus pair 3 the difference between obtained and predicted discrimination of word vowels was higher than it was for text vowels, but at stimulus pair 5 the opposite occurred: there it was higher for the text vowels. The d'-difference scores confirm that there is no difference in the relationship between discrimination and classification as an effect of the naturalness of the vowel stimuli. It could be that there was not enough difference between the two vowel continua to induce differences in perception, in the sense that the vowels from the rapidly read text were perhaps not sufficiently naturally.

Figures 2.5 and 2.6 show that there is not much difference in the results of the two vowel conditions. The data in figure 2.5 were expected to be less categorical than in figure 2.6 and hence the d'-difference scores in figure 2.7 to be much higher for word vowels than for text vowels. This is clearly not the case: the word vowels are not perceived less categorically than the text vowels. Neither figure shows any relationship between observed and predicted discrimination, so it can be concluded that there is no indication of categorical perception for either of the two vowel conditions. Unexpectedly, discrimination performance seems to be worse than predicted by the labelling data. The presumed ordering of the results was: fixed discrimination best, followed by roving discrimination, and then classification (Macmillan et al., 1988; Schouten & Van Hessen, 1992).

The results of a multivariate analysis of variance confirm this interpretation of the results. Fixed independent variables were Task (5 levels), Vowel Condition (2 levels), and Stimuli (7 levels: nested under Vowel Condition). Cell variance was over 19 subjects. There were significant main effects of Task and Stimulus within Vowel Condition (Task F=14.62, p<.001; Stimulus F=5.58, p<.001). Furthermore, there was a significant interaction between Task and Stimulus within Vowel Condition (F=4.56, p<.001). Discrimination and classification performance were not significantly affected by Vowel Condition (F=0.12, p=0.731).

Separate one-way analyses per task with factor Stimulus as independent variable and post-hoc Tukey HSD tests were conducted to test for the significance of classification and discrimination peaks. There were no significant peaks in the discrimination results except for fixed discrimination, word vowel continuum, in which the d'-values of pairs 1 and 2 were significantly higher than those of the other pairs (F=3.811, p=.002). In the classification results there were significant peaks at pair 3 for the word vowels (F=14.39, p<.001), and at pairs 4 and 5 for the text vowels (F=14.24, p<.001). This confirms that there is a peak in the classification results indicating the phoneme boundary, whereas such a peak is completely absent in the discrimination results. Furthermore, discrimination scores are significantly lower than the classification scores at the phoneme boundary location (word vowels F=13.93, p<.001; text vowels F=5.07, p=.003 and F=6.17, p=.001).

In summary, these findings show that classification performance is better than discrimination, which is rather awkward since it indicates that listeners did hear differences between stimuli when classifying them, but could not hear differences between the *same* stimuli during discrimination. It seems as if listeners have applied different perceptual strategies during classification and discrimination. Listeners used a phoneme labelling strategy during classification, but apparently could not assign labels to the stimuli during discrimination. Furthermore, without labelling listeners were incapable of discriminating the stimuli, possibly because the acoustic differences between the stimuli were too small to profit from auditory trace processes. In the next sections, it will be shown that differences among subjects and results from a control experiment confirm that 4I-oddity is a purely psychoacoustic task.

2.3.1 Differences among subjects

A large proportion of the total variance was explained by cell variance (78.5%); this means that there must have been considerable differences, probably in performance level, among the 19 subjects. If our interpretation of the overall results is correct, i.e. if 4I-oddity discrimination is a purely psychoacoustic task that precludes labelling, all subjects should have one thing in common: regardless of their level of performance, none of them should show any relationship between classification and discrimination. A corollary prediction with respect to classification is that, if subjects are graded according to their discrimination performance, this will tell us nothing about classification performance, which should be roughly the same over discrimination quartiles. In order to see whether these predictions were correct, the subjects were divided into quartiles on the basis of the roving 4I-oddity d'-scores.

Figures 2.8 and 2.9 show the discrimination and classification results for the word vowels (the results were the same for the two vowel conditions), obtained from the highest (four subjects) and lowest quartiles (five subjects). There is a marked difference between the discrimination performance of the two subject groups (they were selected on that basis). In figure 2.8 (lowest quartile) the classification results show a peak at the phoneme boundary, with d' scores decreasing to chance level at the extremes of the continuum.

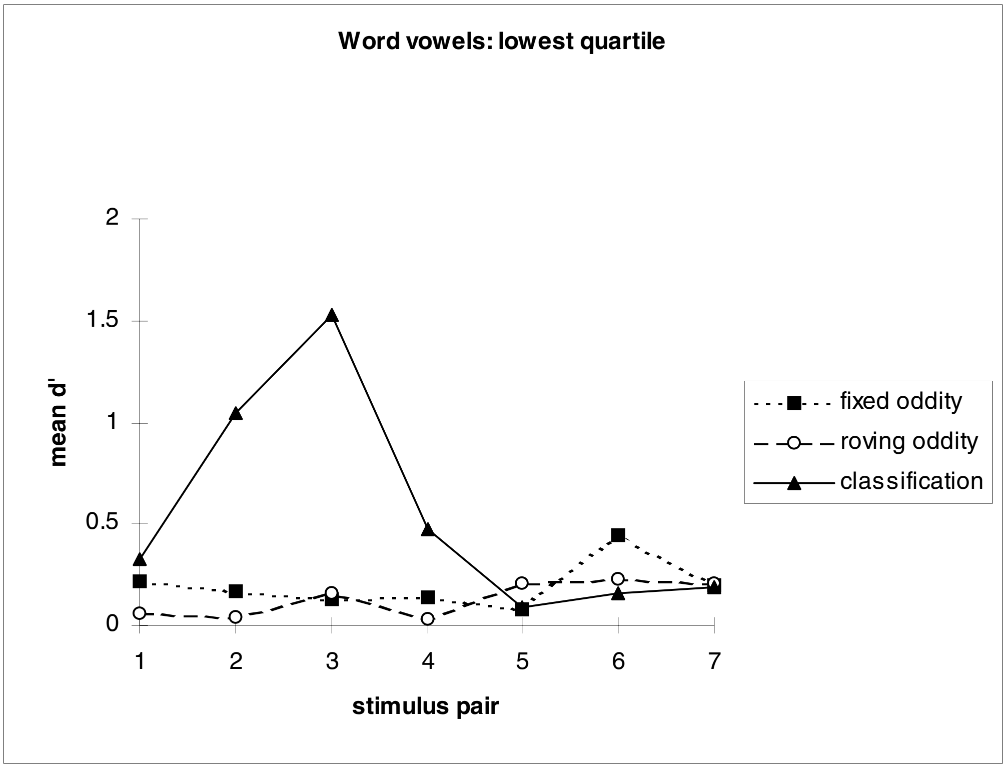


Figure 2.8. Word vowel classification and discrimination of the lowest quartile (5 subjects).

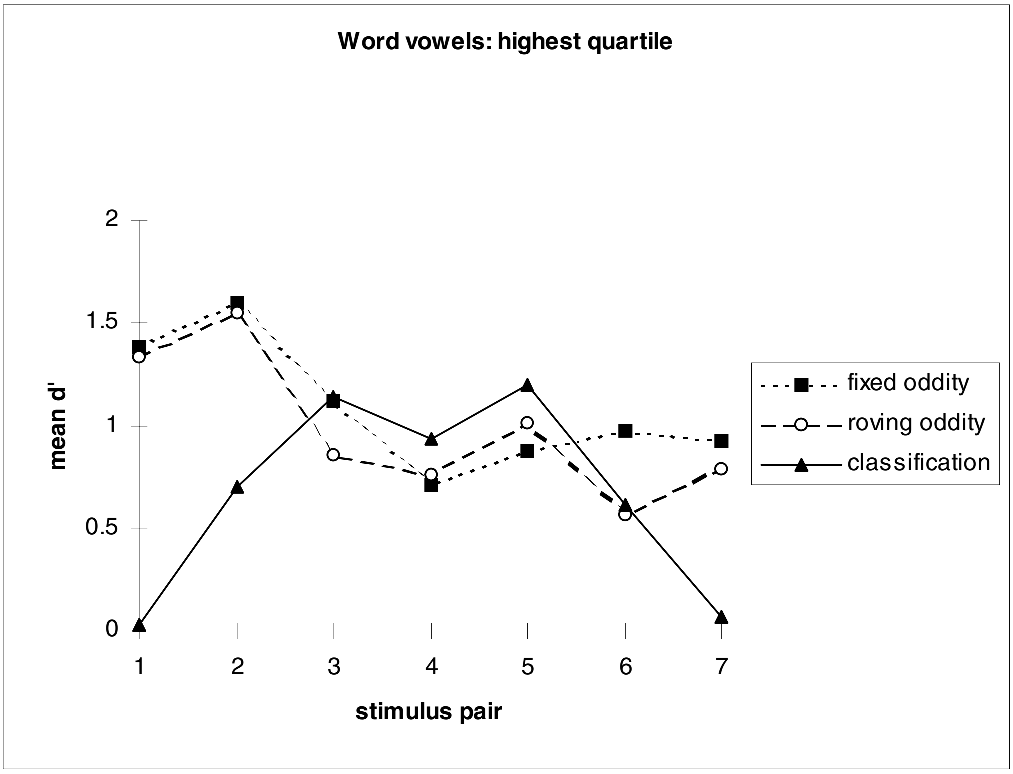


Figure 2.9. Word vowel classification and discrimination of the highest quartile (4 subjects).

The discrimination results are at chance level for all stimuli, indicating that listeners could not detect differences, not even at the phoneme boundary. In figure 2.9, which represents the performance of the subjects in the highest quartile, discrimination is in general as high as, or higher than, classification. The slightly lower d' score at pair 4 is caused by considerable differences in the position of the phoneme boundary (as is often the case with vowel stimuli).

The results for the lowest and highest quartiles show that in neither case does discrimination performance show any relationship with classification performance, which confirms our prediction. This means that, during discrimination, no phonetic information was used: listeners were in the psychoacoustic, or trace, mode. In classification, however, all listeners had to operate in the speech mode and showed the same performance.

These results were confirmed by separate analyses of variance on the data of each task: the effect of the factor Quartiles (4 levels) on discrimination performance was significant (fixed 4I-oddity: $F=9.09$, $p<.001$; roving 4I-oddity: $F=27.60$, $p<.001$), whereas there was no significant effect of Quartiles on classification performance.

The conclusion seems to be almost inescapable: when listening to speech, all subjects perform at roughly the same level, as they would be expected to do in everyday speech situations; when listening in the psychoacoustic mode, however, they show great differences in performance, sorting themselves into good and poor (and intermediate) listeners. This is a common psychoacoustic pattern; several studies have indicated that subjects listening to speech-like stimuli may have different scores if they are operating in a trace mode: some subjects are more ‘analytical’ than others (Best, 1981; Dorman et al., 1977; Foard & Kemler-Nelson, 1984; Repp, 1981; Rosen & Howell, 1981). The fact remains that all our listeners, good or poor, behaved during discrimination as if our stimuli had nothing to do with speech at all.

2.3.2 Control experiment

In the previous section it was shown that discrimination performance for the highest quartile was equal to or better than predicted by the labelling data. However, the highest quartile only contained 4 of the 19 subjects. The discrimination results for the subjects in the other quartiles were worse. This indicates that for these subjects the discrimination task might have been too difficult and therefore that their 4I-oddity results were an artefact of the experiment. To test this hypothesis a control experiment was conducted. In this control experiment the physical distance between the stimuli within trials was changed from one step to two steps along the continuum.

The stimuli used in the control experiment were those from the ‘word vowel continuum’. The subjects were 14 students from the pool of 19 subjects who had participated in the one-step experiment. Unfortunately, 5 students were no longer available for testing. The design consisted of only one task: roving two-step oddity discrimination. In this task 6 stimulus pairs were presented 64 times, following the same procedure as in the previous experiment.

We hypothesised that the larger physical distance between the stimuli would induce a general increase of the discrimination results of all listeners, thus including the subjects in the lowest quartile. Results were not expected to change in the degree of categoricalness, since the results of the highest quartile predict that even if subjects can discriminate between the stimuli, there is no relationship between discrimination and classification performance. The results of two-step discrimination and classification are presented in figure 2.10. The classification results are taken from the previous experiment (minus the data of the five subjects that did not participate in the two-step discrimination task), and recalculated to predict a two-step comparison.

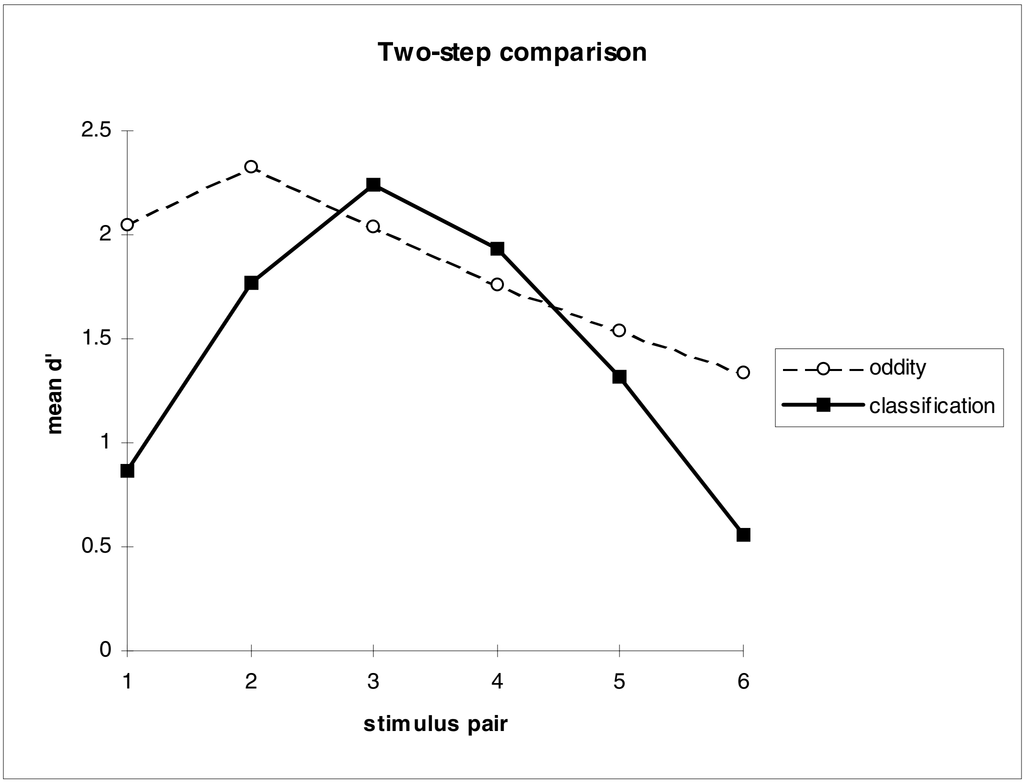


Figure 2.10 Classification and two-step roving discrimination results for the word vowels.

As expected, two-step discrimination generated higher scores than one-step discrimination (see figure 2.5), and the 4I-oddity scores are no longer worse than predicted by the classification data. But again, there is no relation between 4I-oddity discrimination and classification and thus no indication of categorical perception.

The effect of the physical distance between the stimuli was tested with a two-way analysis of variance on the two-step data, combined with the data from the one-step experiment (see figure 2.5, but again the data of the five subjects that did not participate in the control experiment were excluded). The independent variables were Step-size (2 levels) and Task within Step-size (2 levels). Both effects turned out to be significant (Step-size F=142.2, p<.001; Task within Step-size F=6.41, p=.002). A one-way analysis of variance per task with Stimulus as independent variable and post hoc Tukey HSD tests revealed a significant peak in two-step Classification at stimulus pair 3, at the phoneme boundary (F=2.94, p=.02 and F=5.82, p<.001). There was no significant peak in the two-step discrimination results.

Summarising the control condition, it can be concluded that the counterintuitive 4I-oddity discrimination results are not caused by too small a step size between the stimuli. Even with a larger step size, discrimination results show no relationship with the classification data. Again the interpretation of the results must be that different perceptual processes were used for these tasks: an auditory trace strategy during discrimination and phonetic labelling during classification.

2.4 Discussion and conclusion

In a series of experiments the perception of word vowels was compared with that of text vowels. It was expected that the temporal reduction and spectral complexity of the latter would result in a greater dependence on labelling and hence in more categorical perception. However, our results did not confirm this expectation, mainly because there was no relationship between observed and predicted discrimination, and thus no indication of categorical perception for either of the two vowel conditions.

Discrimination performance was poor and fell generally below that predicted on the basis of classification. This paradoxical finding (paradoxical because discrimination performance has nearly always been better than predicted - see, e.g., Macmillan et al., 1988; Pastore, 1987; Repp, 1984; Schouten & Van Hessen, 1992) could not be explained away as being due to the unprecedented difficulty of the task: a greater physical spacing between stimuli in a discrimination trial did not make any significant difference.

A first attempt to account for the traditional difference (discrimination better than predicted by the classification data) was made in the dual-process model for the discrimination of speech stimuli by Fujisaki and Kawashima (1971). This model explicitly distinguishes between categorical phonemic judgements and judgements based on auditory memory for acoustic stimulus attributes. The authors propose that two perceptual modes are active simultaneously (or in rapid sequence). One of them is strictly categorical and represents phonetic classification and the associated verbal short-term memory. The other mode is not categorical and represents processes common to all auditory perception. The results of any particular speech discrimination experiment are assumed to reflect a mixture of both components. The part of performance that can be predicted from labelling probabilities is attributed to categorical judgements, whereas the remainder (the deviation from ideal categorical perception) is assigned to comparison of acoustic stimulus properties (Repp, 1984).

The selection of a discrimination paradigm was led by the idea that a discrimination task in which trace coding would not be excluded in advance was needed to assess categorical perception. If a discrimination task is used that *prevents* a direct comparison between successive stimuli, listeners are *forced* to use a phonetic labelling strategy for discrimination and results will *inevitably* be highly categorical. The task was even more successful than expected in stimulating a trace coding strategy: the results show that there was no categorical perception at all. Moreover, discrimination performance was lower than predicted by classification. This was especially true of the lowest quartile, whose classification results were quite normal, but whose discrimination results were at chance level for all stimuli, indicating that listeners could not detect differences, not even at the phoneme boundary. The discrimination performance of the subjects in the highest quartile, however, was in general as high as, or higher than, classification. In neither case did discrimination performance show any relationship with classification performance. Similar results were found for all subjects in the control experiment with a larger physical distance, which made discrimination easier but did not lead to a stronger relationship between discrimination and classification results. This means that, during discrimination, no phonetic information was used: listeners were in the psychoacoustic, or trace, mode. In classification, however, all listeners had to operate in the phonetic mode and showed the same behaviour. Why is it that 4I-oddity puts subjects into the psychoacoustic mode, whereas all other speech discrimination tasks that have been used up to now produce at least a mixture of the two modes, so that results always turn out categorical to some extent?

It is suspected that this is due to the nature of the oddity task, which is less close to the traditional 2IFC task than was expected, and closer to 4IAX. The crucial difference here is whether a subject's decision is influenced by a criterion, or bias, that is external to the stimuli. In 2IFC, in which a subject has to indicate the order of two stimuli, this can only be done with reference to criteria that are external to the stimuli, such as phoneme boundaries or categories such as "high" vs. "low" in psychoacoustic experiments. However, the oddity results were more like those that may be expected from 4IAX, in which subjects are presented with two pairs of stimuli and have to decide which pair contains the odd one out. The 4IAX task does not refer subjects to criteria that are external to the stimuli: subjects hear two differences and have to decide which one is the greater one, a decision that can only be taken on the basis of stimulus information and not on the basis of a criterion along a scale. This is why 4IAX is generally regarded as bias-free in psychophysics.

Subjects' comments, and our own experience when performing the oddity task, confirm the similarity between 4I-oddity and 4IAX, at least with vowel stimuli: it soon became apparent to subjects that attempts to label the stimuli did not work, so they gave up the attempt and apparently applied a strategy in which differences between two pairs of stimuli were used in order to reach a decision. As a result, the discrimination experiment turned out to be one in which phoneme categories did not play any role whatsoever. This was unexpected: even though Heller and Trahiotis (1995) have suggested that subjects might treat 4I-oddity as a variant of 4IAX, it still was not foreseen that there would be a complete absence of phonetic information in the discrimination response. 4IAX may be bias-free, but phoneme labels were expected to be inevitable: a human subject should not be able to treat speech as if it does not consist of phonemes or words. This is wrong: it *is* possible to get subjects completely into the psychoacoustic trace mode when they have to discriminate speech sounds.

In normal, everyday speech perception, we perceive categorically. It has always been assumed that this mental process can be investigated by looking at the relationship between two psychophysical tasks: classification and discrimination. The present results suggest that this assumption rests crucially on the use of biased discrimination tasks, in which listeners are compelled to use subjective decision criteria. When a bias-free discrimination task is used, no resemblance between phoneme classification and discrimination is found at all. This indicates that categorical perception is highly task dependent. In the next chapter, the hypothesis that the degree of categorical perception varies as an effect of the discrimination task and of the duration of the inter-stimulus interval was explicitly tested.

Chapter 3

The effect of task and inter-stimulus interval

3.1 Introduction

The traditional classification test in categorical perception research is a forced choice between two or more phoneme labels. For testing speech discrimination, several different paradigms can be used. In the preceding chapter we described results in a 4I-oddity discrimination task, which had never been used before to test categorical perception. The choice of this discrimination task was justified by criticism of discrimination tasks such as ABX, e.g. by Massaro and Cohen (1983) and Kluender (1994), who argued that ABX discrimination induces phonetic labelling strategies and hence produces results that are usually attributed to categorical perception.

In 4I-oddity, we expected listeners to use phoneme labelling, but also to make direct auditory comparisons between the speech stimuli in a trial. Discrimination performance turned out to be completely unrelated to classification performance: classification results showed a marked peak at the phoneme boundary, but the shape of the observed 4I-oddity discrimination function was flat and d' was even lower than predicted at the phoneme boundary location. This lack of any relationship between the discrimination and classification results is counterintuitive to anyone who has ever spent any time comparing classification and discrimination in a categorical-perception context. Not only was there no relationship at all between classification and discrimination, but subjects did not even manage to detect any differences between stimuli that, during classification, they had, fairly consistently, given different phonetic labels.

The 4I-oddity results indicate that there is something about this discrimination task that prevents subjects from using phonetic information. If this is true, it could point to two different perceptual strategies: an auditory comparison of stimulus information during discrimination and phonetic labelling during classification. In other words, during discrimination listeners are in a ‘psychoacoustic mode’ and during classification in a ‘speech categorisation mode’. However, before allowing ourselves to draw such a far-reaching conclusion we decided to investigate the effect of two procedural variables on the previous vowel results: 1) the effect of the discrimination task; 2) the effect of the duration of the inter-stimulus interval.

In the present chapter, three experiments will be described. In Experiment I, we compared the relationship between classification and performance on three different discrimination tasks: 2IFC, 4IAX and 4I-oddity. Experiment II was conducted with the same discrimination tasks as in experiment I, but this time a longer inter-stimulus interval was used to test the effect of inter-stimulus-interval duration on categorical perception. In Experiment III, we added AX and AXB discrimination to the task design of experiment I, since task turned out to have a relatively large effect on the degree of categorical perception.

3.2 Experiment I: 2IFC, 4IAX and 4I-oddity discrimination

As already noted, the reason for choosing the four-interval oddity task for the experiment in chapter 2 was to avoid an exclusive reliance on phoneme labelling processes during discrimination. The results seem to indicate that we succeeded too well: the discrimination results deviate from the classification results in a way suggestive of an exclusive reliance on psychoacoustic trace coding. In order to see if our oddity results do reflect an auditory trace strategy we decided to compare them with performance in other discrimination tasks, notably a discrimination paradigm that tends to encourage auditory-trace processing and one that tends to encourage labelling.

A task in which phonetic labelling processes are known to play an important role is 2IFC discrimination. In this “two-interval, two-alternative forced choice” paradigm, the stimuli are always different and the subject has to determine the order in which they are presented. To indicate this stimulus order, the response fields are labelled “ie-oe” and “oe-ie” (/u-i/ and /i-u/). The explicit reference to phoneme categories in the instructions is assumed to encourage phoneme labelling processes. If this assumption is correct, we should obtain the “regular” relationship between classification and 2IFC discrimination, in which discrimination can be predicted to some extent by, and does not fall below, classification.

Unlike 2IFC, the 4IAX discrimination task is said to be sensitive to acoustic differences between speech stimuli (Pisoni, 1975). The trials in this task consist of eight combinations: ABAA, BAAA, AAAB, AABA, and BABB, ABBB, BBBA, BBAB. In each trial the silent interval between stimuli 2 and 3 is longer than between the other stimuli, so that listeners hear two pairs within one trial. They have to decide which pair contained identical stimuli, pair 1 or pair 2. It is assumed that listeners first determine the difference between the stimuli within the pairs and then determine which of the two differences is the smaller one. If both 4IAX and 4-interval oddity draw upon a comparison of the auditory traces of the stimuli we expect their results to be similar. This means that we predict that the 4IAX discrimination results will be unrelated to the prediction from the classification data.

In short, we expect a high degree of categorical perception for vowels tested with 2IFC discrimination, and noncategorical results for vowels tested with 4IAX and 4-interval oddity discrimination.

3.2.1 Method

3.2.1.1 Stimulus material

The stimuli used in this study were identical to the ‘word vowel continuum’ described in section 2.2.1 of chapter 2. The stimuli were taken from a seven-step continuum between the vowels in the words /pup/ and /pip/, which were read aloud in a list of isolated words by a male, native speaker of Dutch.

3.2.1.2 Subjects

In each test, 14 subjects were used. They were 14 students from the pool of 19 subjects who had participated in the experiments in chapter 2. The subjects had no known hearing deficits, they were all native speakers of Dutch and were paid for their participation.

3.2.1.3 Design

Each subject took part in 3 tests: one classification test and two discrimination tests. Discrimination was tested with 2IFC and 4IAX discrimination. For the 4I-oddity results we used the data from chapter 2, but this time only from the 14 subjects that participated in the two new discrimination tasks.

In the discrimination test, stimuli were presented in fixed and roving contexts. In the fixed context, stimuli were presented in 7 blocks, one for each stimulus pair. Each block contained 64 trials with the same two stimuli. In the roving context, stimuli in each trial were drawn at random from the total set of stimuli. In total 7 x 64 trials were presented. There was no response-time limit in discrimination or in classification.

In 2IFC discrimination, trials consisted of two different stimuli, either AB or BA. The subjects responded by indicating the order in which the stimuli were presented. The inter-stimulus interval was 200 ms.

In 4IAX discrimination each trial consisted of two stimulus pairs: AB-AA, BA-AA, AA-BA, AA-AB, BA-BB, AB-BB, BB-AB, BB-BA. These eight combinations were presented 8 times in random order. The listeners had to indicate which of the two pairs within each trial contained identical stimuli. The ISI was 200 ms and the interval between pairs was 500 ms.

The trials in 4I-oddity consisted of AABA and ABAA combinations. Subjects had to indicate if the oddball was in the second or in the third interval. The inter-stimulus interval was 200 ms.

Classification involved a forced choice between two alternatives, the vowels /u/ and /i/. Each stimulus (8) had to be identified 64 times in a random order.

3.2.1.4 Procedure

The stimuli were presented to the subjects over headphones in a sound-treated booth. Each discrimination task was preceded by a training session, consisting of 56 trials. Training was intended to familiarise subjects with their task. In the fixed discrimination context, there were

10 extra trials at the beginning of each block, which were considered practice and were not included in the data analysis.

The subjects responded by mouse-clicking on one of two response fields on a computer screen. The labels of these response fields varied as a function of the discrimination task. In 2IFC discrimination they were labelled “oe-ie” and “ie-oe” (“oe” = /u/ and “ie” = /i/), in 4IAX the response labels were “pair 1” and “pair 2” and in 4I-oddity they were labelled “2” or “3”. After answering, visual feedback (250 ms) of the correct answer was given so that the subject was able to judge and possibly improve his or her performance.

In classification, one stimulus was played on each trial, and the subject had to identify it by mouse-clicking on a response field labelled “oe” or “ie”. No feedback about correct responses was given.

3.2.2 Results 2IFC, 4IAX and 4I-oddity

The results are presented in figure 3.1. The data in the figure represent the average classification and discrimination of the 14 subjects’ individual d' scores. The numbers (n) along the abscissa refer to stimulus pairs, consisting of stimuli (n) and (n+1). Stimuli in pair 1 resemble /u/ and stimuli in pair 7 sound like /i/. The results of the classification test are presented as predicted discrimination scores. The d' scores were calculated in accordance with the standard assumptions about optimum decision strategies (Macmillan & Creelman, 1991). The values of p(H) and p(FA) were forcibly limited to the 0.99 to 0.01 range, which meant that the maximum d'-value that could be obtained was 3.29 for 2IFC and 4I-oddity. The maximum d'-value for 4IAX was 5.14, and for classification it was 4.65.

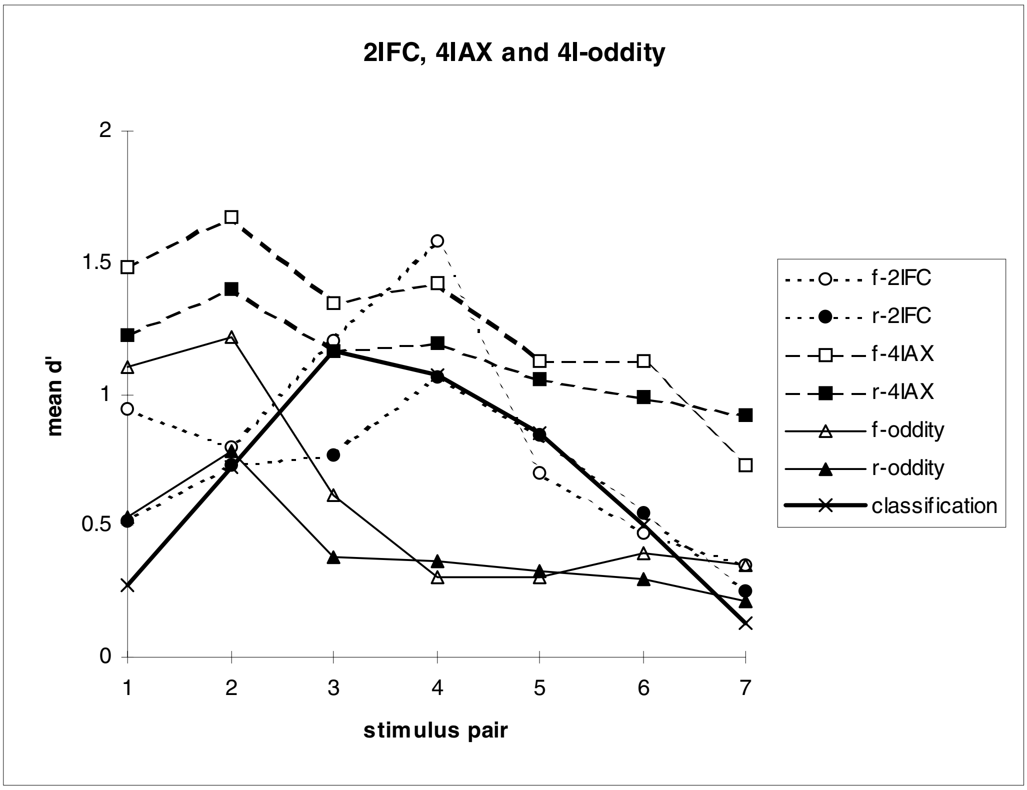


Figure 3.1. Classification and 2IFC, 4IAX and 4I-oddity discrimination results for the word vowels.

As can be seen in figure 3.1, 4IAX and 4I-oddity discrimination performance shows no close relationship with the classification data. The results of 2IFC discrimination, however, seem to be more closely related to classification (the lower d'-value at stimulus pair 3 in roving 2IFC is caused by the atypically low score of one subject). As expected, fixed discrimination is always better than roving discrimination. Furthermore, the peak in fixed 2IFC indicates that the small stimulus range reduced context variance and consequently that these results show a marked labelling effect.

A two-way analysis of variance with factors Task (7 levels) and Stimulus (7 levels) revealed main effects for both factors: Task, F=18.52, p<.001, Stimulus, F=8.34, p<.001. There was also a Task by Stimulus interaction, F=1.70, p=.007. Separate post hoc Tukey HSD with Task showed that the ordering of the d'-values, from high to low was as follows: 4IAX 2IFC, classification and 4I-oddity. Post hoc tests for the factor Stimulus revealed that d'-scores for stimulus pairs 2, 3, and 4 were significantly higher than for the other pairs.

To test for the significance of the classification and discrimination peaks and valleys, one-way analyses of variance and post-hoc Tukey HSD tests were carried out on each separate task with Stimuli as independent variable. This was done mainly to test the significance of the classification and discrimination peaks. Stimuli were defined to constitute a peak if they led to significantly higher d' values than did the majority of the other stimuli. For classification there was a significant peak at stimulus pairs 3 and 4 (F=6.17, p<.001). For 2IFC there was a significant peak at pair 4 (F=2.68, p=.019). At pair 4 there was no significant difference between the d'-scores for classification and fixed and roving 2IFC discrimination, which confirms that there was a close relationship between classification and observed 2IFC discrimination. There were no significant peaks in the 4IAX and 4I-oddity results. At pairs 3 and 4, the d'-scores for 4I-oddity were significantly lower than the d'-scores for the other tasks. 4IAX performance was significantly higher than performance on the other tasks at pairs 1, 2, 6 and 7.

In the previous chapter we saw that auditory-trace discrimination varied among subjects. To gain more insight into the strategies used by the listeners, we again divided the results into quartiles on the basis of the 4I-oddity results. The lowest and highest quartiles contained 4 subjects each. In figures 3.2 to 3.5 the average d' scores are presented of subjects in the lowest quartile and in the highest quartile. Since the d' functions in these figures are plotted closely together, which causes these results to be more difficult to interpret than the overall results, separate figures with cumulative d' are also presented.

There were two expectations concerning the difference in performance between these quartiles. Firstly, it was predicted that there would be significant differences between the performance of listeners in the lowest and highest quartiles on both 4-interval tasks. This would indicate that the same auditory-trace strategy is used in 4I-oddity as in 4IAX. Secondly, it was expected that listeners in the lowest and highest quartiles would not differ in their 2IFC and classification performance, because these tasks were assumed to be mediated by the same perceptual process, viz. phoneme labelling.

A one-way analysis with factor Quartiles (4 levels) showed a significant effect, F=26.82, p<.001. Separate analyses of variance and post-hoc Tukey HSD tests on the data per task showed that, as predicted, performance on classification was not significantly different. Performance on all discrimination tasks turned out to be worse for the lowest quartile than for the highest quartile at the p<.001 level. Contrary to our predictions, 2IFC results were significantly different as an effect of factor Quartiles, F=5.268, p=.002. It appears that the listeners in the highest quartile were not only better at hearing small acoustic differences between stimuli, but also at labelling them.

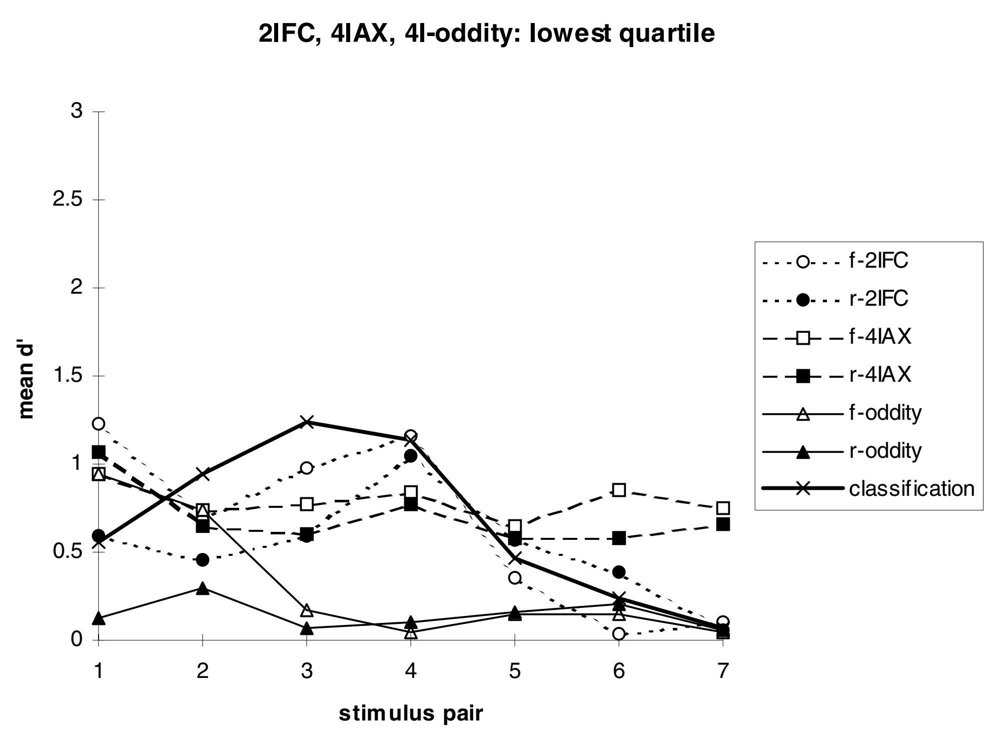


Figure 3.2 Vowel classification and 2IFC, 4IAX, 4I-oddity discrimination of the lowest quartile (4 subjects).

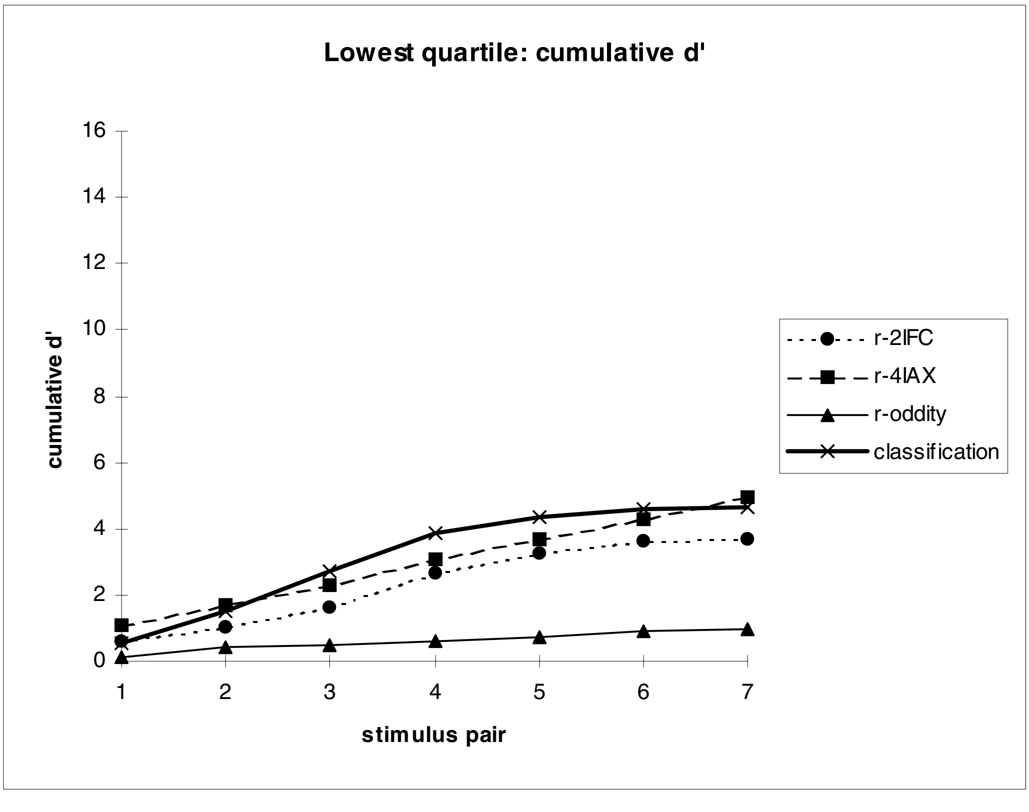


Figure 3.3 Cumulative results of the lowest quartile.

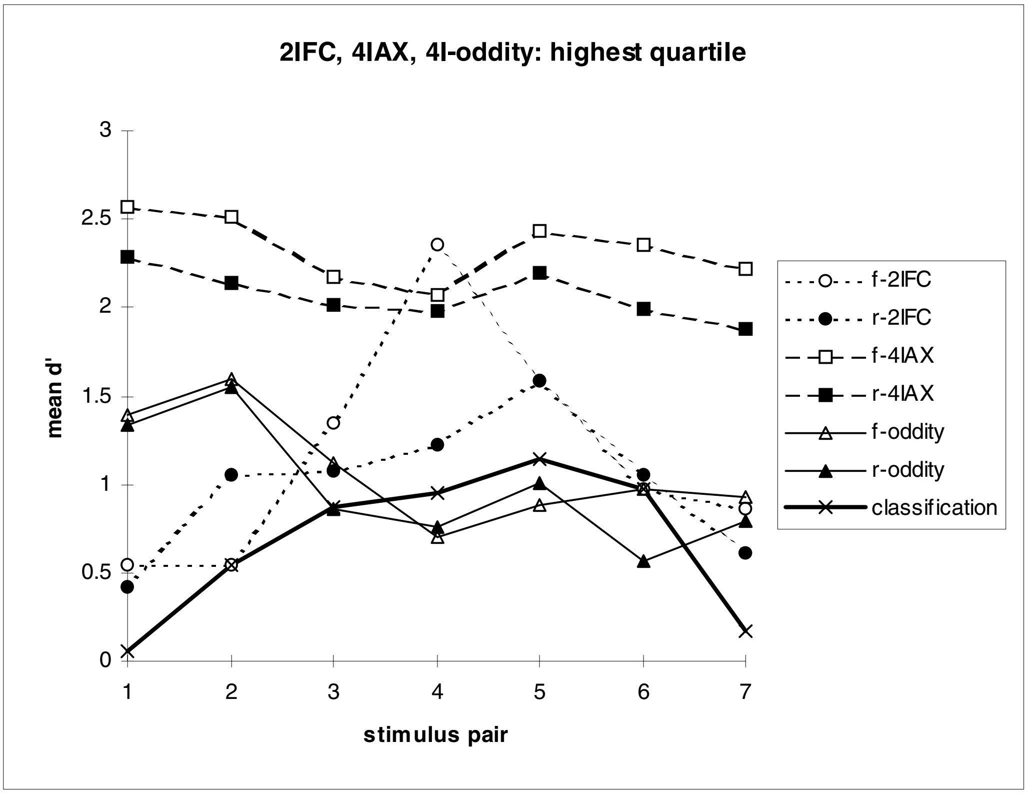


Figure 3.4 Vowel classification and 2IFC, 4IAX, 4I-oddity discrimination of the highest quartile (4 subjects).

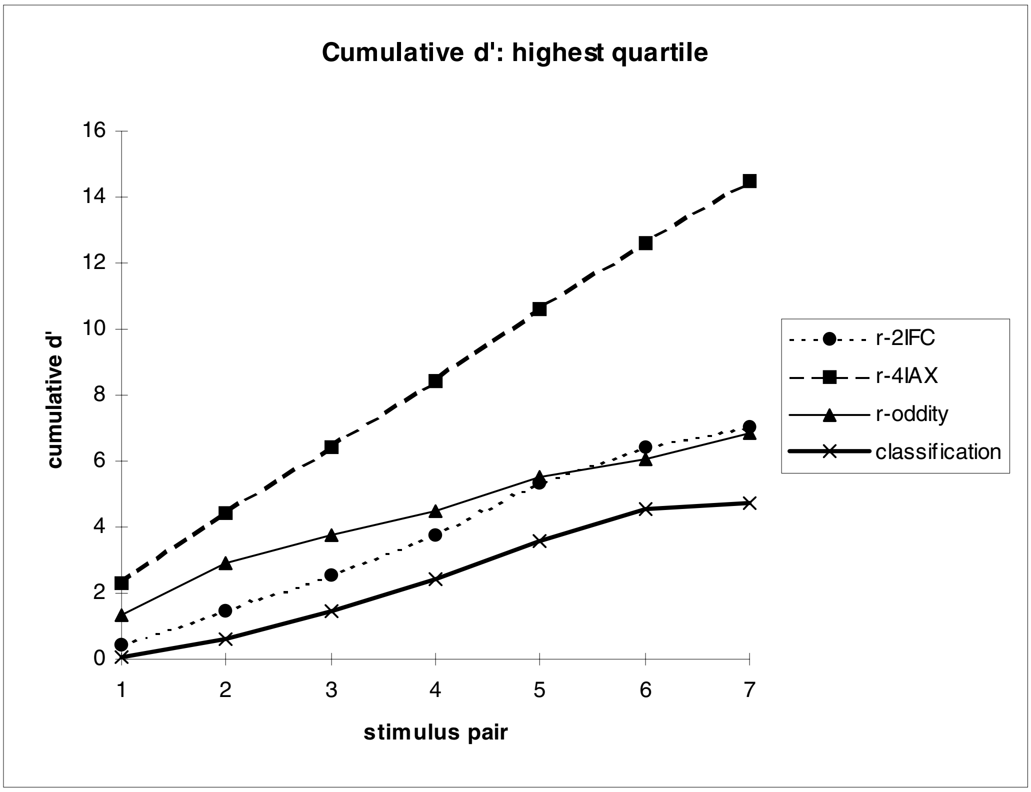


Figure 3.5 Cumulative results of the highest quartile.

This implies that the performance differences may be due to more general cognitive processes, such as motivation and concentration. However, it has to be noted that the difference between quartiles was larger for the 4I-tasks than for 2IFC: cumulative d' for 4IAX was 2.94 times as high for the highest quartile as for the lowest quartile. For 4I-oddity it was 6.91 times as high and for 2IFC 1.89 times.

If we look at the individual results, it turns out that 2 of the 14 listeners showed a peak in 4IAX, but also much better than predicted within-category discrimination. In 4I-oddity, none of the listeners had a peaked discrimination function. This indicates that, for some listeners, the decision strategies for 4I-oddity and 4IAX were not as similar as we expected. The main difference between the two tasks was the interval between pairs in 4IAX, which was 500 ms. The results suggest that this relatively long interval may increase the memory load of the task in a way that encourages a subgroup of the listeners to use a labelling strategy. Unusual sensitivity to phonetic distinctions by some listeners has been reported before (e.g. MacKain, Best & Strange, 1981).

To summarise, the present results show different degrees of categorical perception for the same vowel stimuli as a function of the discrimination task. It is evident that 2IFC discrimination is more closely related to classification performance than is discrimination with the 4-interval tasks. Therefore, vowel perception is more categorical with 2IFC discrimination, and much less categorical with 4IAX and 4I-oddity discrimination. This implies that different perceptual strategies were used in the different discrimination tasks. The 2IFC results suggest that subjects used labelling, whereas the 4IAX and 4I-oddity results imply that an auditory-trace comparison was the basic strategy used by listeners. The 4IAX results further indicate that the degree of categorical perception also differs among listeners, since a small subgroup showed a labelling peak in their discrimination function whereas others did not.

3.3 Experiment II: Duration of the temporal interval between stimuli

In reports that criticise the categorical perception assumption it is often argued that the close relationship between discrimination and classification performance is caused by the high memory load in discrimination (e.g. Cutting, Rosner & Foard, 1976; Pisoni, 1973). The listener has to compare two or more speech stimuli and uses the phoneme representations in long term memory to accomplish this task, especially when the inter-stimulus interval (ISI) is relatively long.

In experiment I we had, on the basis of Massaro’s studies on auditory storage time, decided to use an interval of 200 ms, in which trace information should be available for direct auditory comparison (Massaro, 1972a, 1972b, 1974). His results have shown that after approximately 200 ms the auditory trace of a sound signal has faded. However, as a result of this relatively short interval, listeners may have been discouraged from using a labelling strategy during 4I-oddity discrimination, because there may not have been enough time between the successive stimuli to synthesise the acoustic information into phonetic percepts. This interpretation is suggested by the 4IAX data that contain labelling effects and in which the interval between pairs was 500 ms instead of 200 ms. In line with the 4IAX results, Cowan & Morse (1986), Pisoni (1973), and Van Hessen & Schouten (1992), have shown that discrimination performance based on labelling reaches a maximum after between 500 ms and 1000 ms; if this is correct, it could explain the lack of any relationship between classification and 4I-oddity discrimination in our data.

In addition, the low overall 4I-oddity discrimination scores could mean that the interval of 200 ms was also too short to extract the acoustic information from the four stimuli in a trial. Listeners may need more processing time in the 4I-oddity task than they do in the masking task Massaro (1972a, 1972b, 1974) used in his experiments. In other words, auditory analysis and comparison of four successive speech sounds may require more time than 200 ms, and the arrival of new stimuli may disrupt processing of the previous stimuli.

In the present experiment we expanded the inter-stimulus interval from 200 ms to 500 ms. For 4I-oddity we predicted that discrimination performance would increase independently of classification if this longer ISI is enough to allow listeners to use trace information for discrimination as a result of reduced interference between the four stimuli within a trial. Moreover, the listener will have more time to assign labels to the stimuli (Van Hessen & Schouten, 1992), and observed discrimination may therefore increase differentially as a function of classification (predicted discrimination). This means that stimuli with different phoneme labels may now be easier to discriminate than stimuli within one phoneme category, and thus that perception may be more categorical.

The effect of the longer ISI was also tested with 2IFC and 4IAX discrimination. For 2IFC we expected discrimination to show an unchanged phoneme labelling effect and to improve with increasing ISI, as was shown in Van Hessen and Schouten (1992). For 4IAX, we hypothesised discrimination to improve in a way that may show more effect of the labelling strategy compared to the baseline condition.

3.3.1 Method

The stimulus material was again the word vowel continuum and the 14 subjects from experiment I also participated in the present experiment. The test procedure was identical to the one in experiment I (section 3.2.1.4).

3.3.1.1 Design

Experiment II consisted of three tests: 2IFC, 4IAX and 4I-oddity discrimination. The stimuli were only presented in roving context. The design was identical to that of experiment I, except for the duration of inter-stimulus interval. In experiment I, ISI had been 200 ms and this was changed to 500 ms. In 4IAX discrimination, the inter-pair interval was also changed: from 500 ms to 1000 ms.

3.3.2 Results long-ISI condition

Figure 3.6 displays the results of roving discrimination and classification in the long-ISI condition. The classification data are taken from experiment I in this chapter. The figure shows that, compared to 4IAX and 4I-oddity discrimination, the 2IFC discrimination results are best predicted by the classification data. There is a broad peak in 2IFC, which coincides fairly well with the peak in classification. The 4I-oddity performance is still worse than predicted by the labelling data, except for the endpoint stimuli, and thus there is no relationship between these discrimination results and classification. The 4IAX results are overall much better than predicted and the shape of the function shows no relation with the shape of the classification function. The present results are very similar to the results of experiment I in which the same tasks were used but with a shorter ISI (see figure 3.1). This similarity indicates that the degree of categorical perception has not changed as an effect of the longer ISI.

The effect of ISI was tested with a three-way analysis of variance on the data from the current experiment, combined with the data from experiment I. The independent variables were ISI (2 levels), Task nested under ISI (3 levels) and Stimulus (7 levels). The classification data were excluded from this statistical analysis since the effect of ISI does not play a role in classification. Cell variance was over 14 subjects.

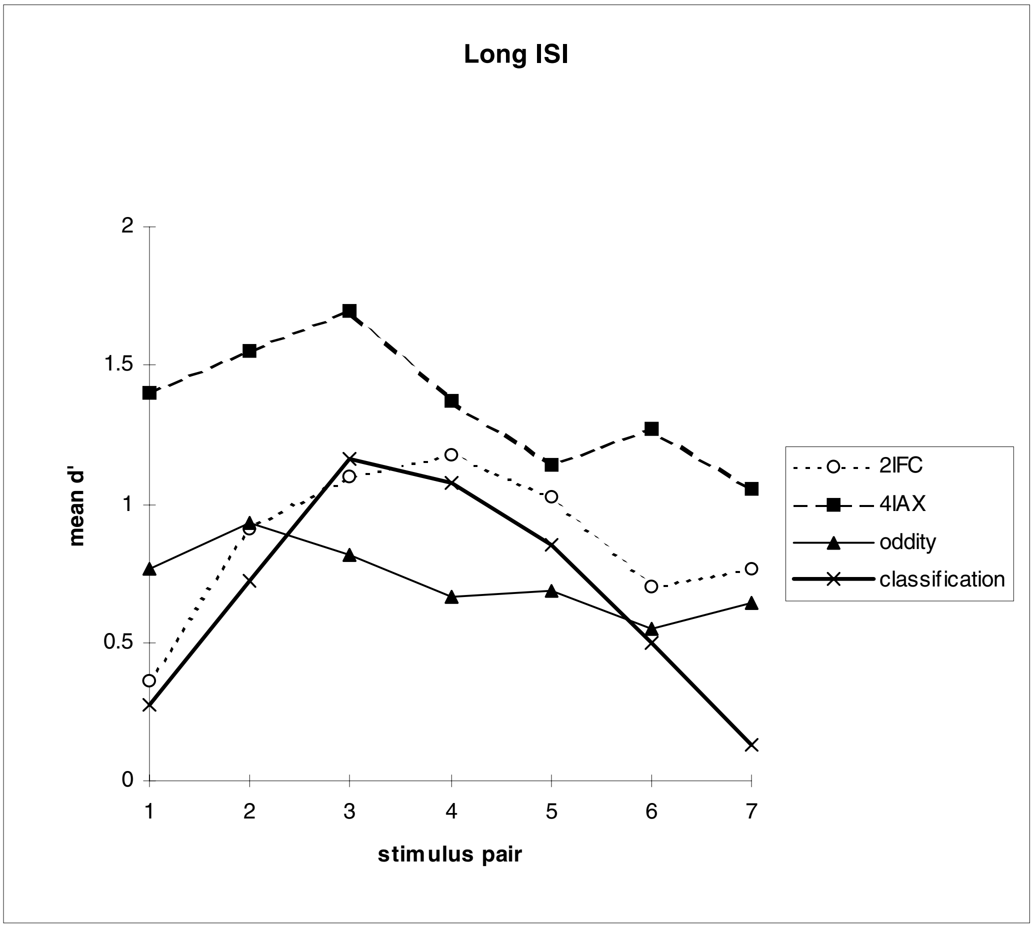


Figure 3.6. Long ISI: classification and roving 2IFC, 4IAX and 4I-oddity discrimination results for the word vowels.

The statistical analysis revealed that there were main effects of ISI, Task nested under ISI, and Stimulus (F= 21.15, p<.001; F=26.23, p<.001; and F=3.70, p=.001, respectively). The main effect of ISI means that the long-ISI d'-values were significantly higher than the short-ISI ones, which indicates that the longer ISI has facilitated discrimination. The effect of the longer ISI was largest for 4I-oddity, which showed an increase of 74%. 4IAX improved by 27% and 2IFC was 30% better.

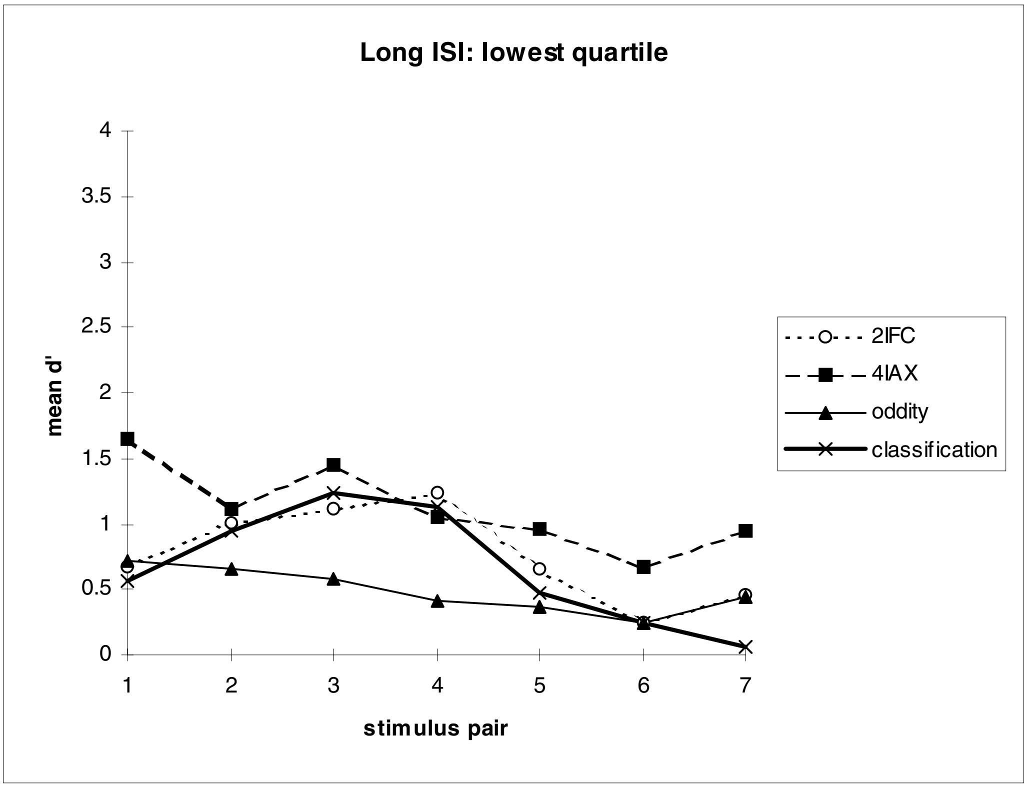


Figure 3.7 Long-ISI condition: vowel classification and 2IFC, 4IAX, and 4I-oddity results of the lowest quartile (4 subjects).

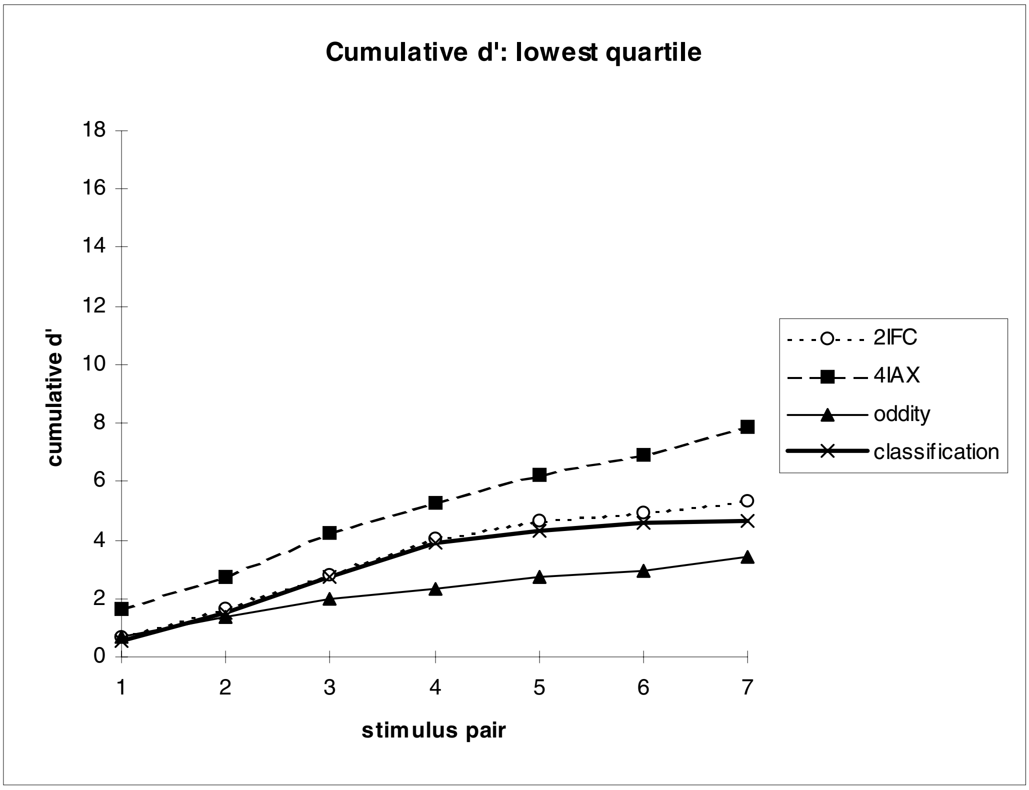


Figure 3.8 Long-ISI condition: cumulative d' of the lowest quartile.

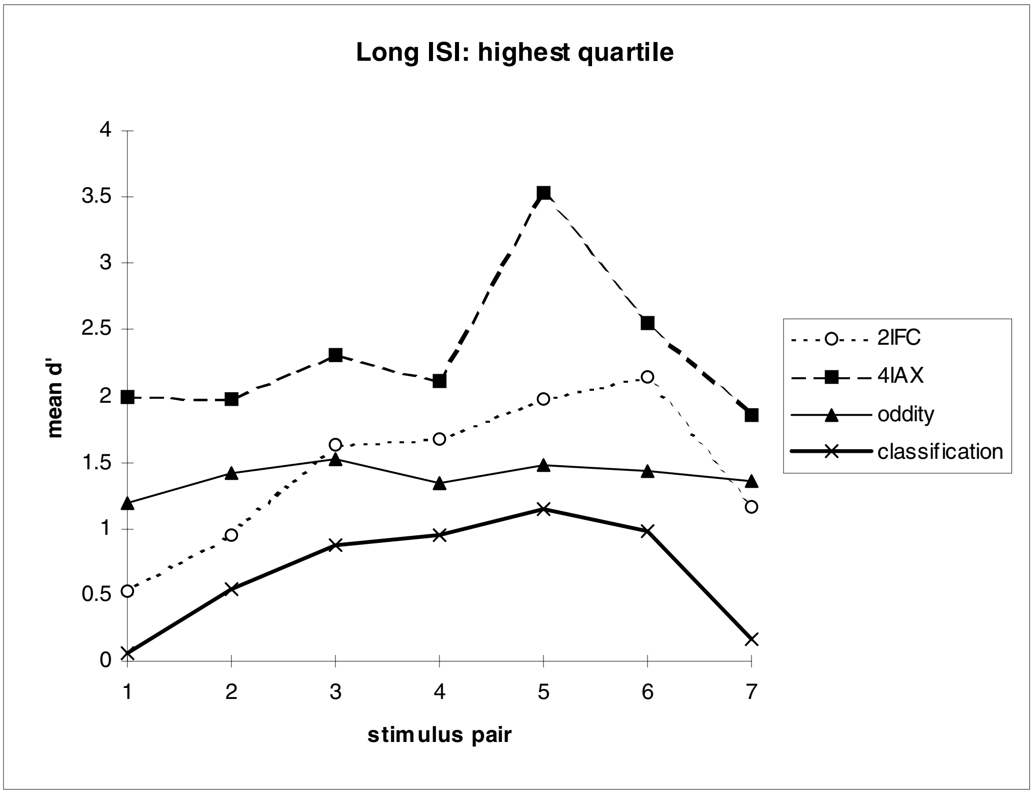


Figure 3.9 Long-ISI: vowel classification and 2IFC, 4IAX, and 4I-oddity results of the highest quartile (4 subjects).

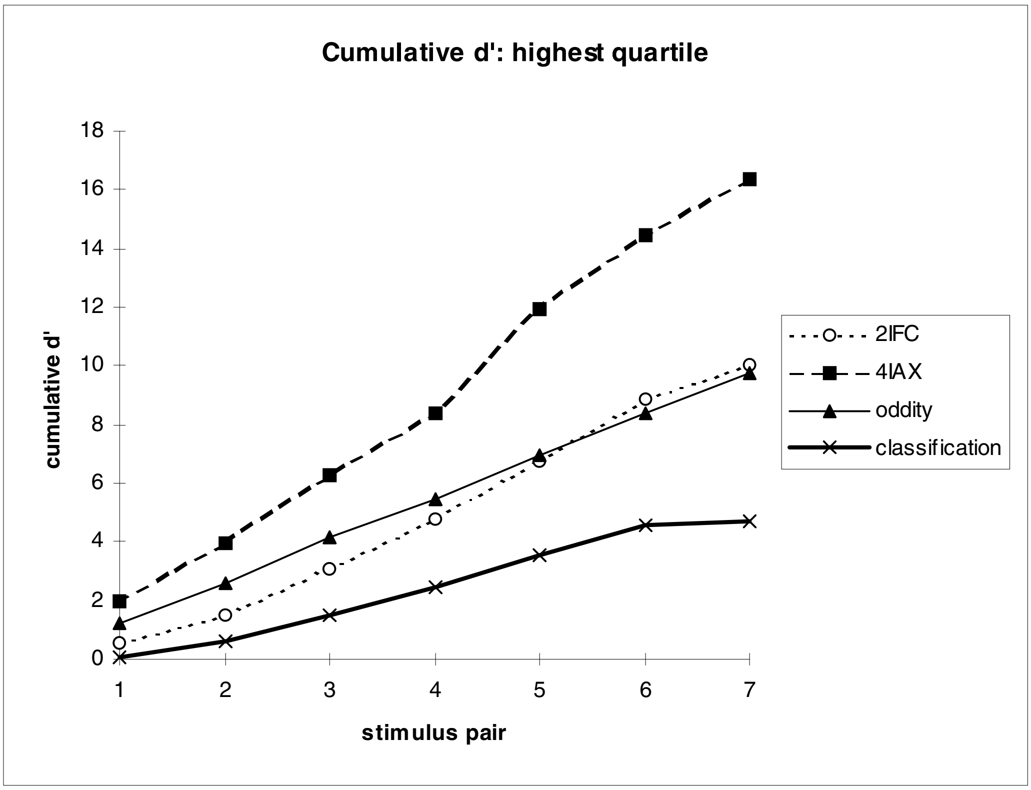


Figure 3.10 Long-ISI: cumulative d' of the highest quartile.

The large improvement in the oddity results indicates that listeners did need more processing time than the 200 ms in the baseline condition and that the low scores were partly caused by the disruption of processing by the rapid presentation of 4 successive speech stimuli.

Separate analyses of peaks and valleys in the functions revealed that the peak in 2IFC was not significant (F=1.96, p=.08), probably due to the high d' value at stimulus pair 7, which was also the only pair at which 2IFC performance was significantly higher than classification performance, F=4.64, p=.006. The statistical analysis confirms what was concluded on the basis of the patterns: the longer ISI has not changed the degree of categoricalness. However, figure 3.9 shows that the perception of some subjects is more categorical in the long-ISI condition than in the short-ISI condition (see figure 3.4). In this figure, which displays the average results of 4 subjects in the highest quartile, there is a peak in 4IAX discrimination, which was absent in the shorter ISI results. On the other hand, in figure 3.7 with the performance of the subjects in the lowest quartile, it can be seen that long-ISI discrimination is similar to short-ISI discrimination (see figure 3.2). If we look at the individual results, it turns out that 5 of the 14 listeners showed a peak in the 4IAX results, which is suggestive of a phoneme-boundary effect. Still, listeners' performance is always better than predicted for the within-category stimuli. In 4I-oddity, none of the listeners had a peaked discrimination function.

A separate one-way analysis of variance for Quartiles (4 levels) on the long ISI data revealed a significant effect, F=29.02, p<.001. Post-hoc Tukey HSD tests on the data per task showed that the results of the highest quartile were significantly higher than those for the other quartiles on all tasks except for classification, which is exactly what we found for the shorter ISI condition. Again the difference between quartiles was larger for the 4I-tasks than for 2IFC: cumulative d' for 4IAX, the highest quartile, is 2.94 times as high as the lowest quartile.For 4I-oddity this was 2.86 times and for 2IFC discrimination 1.89. In the baseline condition, cumulative 4I-oddity d' was 6.91 times higher for the highest quartile, which confirms the overall finding that it was mainly oddity discrimination that benefited from the longer ISI, especially with the subjects in the lowest quartile.

It has to be concluded that there is an effect of inter-stimulus interval on discrimination, but this effect is rather small and mainly affects overall performance. In general, there is no effect of ISI on the degree of categorical perception. Only in the 4IAX task did the longer ISI change the perceptual strategy of 2 of the 14 listeners: 5 subjects had a peak in 4IAX, which was only present in the results of 2 listeners with the shorter ISI. Nevertheless, average 4IAX performance is not at all related to classification performance and hence perception is not categorical.

The scores in 2IFC discrimination have increased, but still agree quite well with classification, indicating that perception is fairly categorical. There is a marked increase of 4I-oddity scores, but again no relationship between observed and predicted discrimination. None of the individual subjects showed a discrimination peak. It can be concluded that the low oddity discrimination results in the shorter ISI condition were partly caused by interference of the 4 successive stimuli. However, the lack of a relationship between discrimination and classification performance was not caused by too short an inter-stimulus interval between the stimuli, but by the psychoacoustic nature of the task, which precludes phoneme labelling.

3.4 Experiment III: AX and AXB discrimination

In experiments I and II we have seen that the degree of categorical perception changes dramatically as a function of the discrimination task, and to a lesser extent as a function of inter-stimulus interval. Since the effect of task appears to be rather large, we added two tasks to the design of experiment I: AX discrimination and AXB discrimination. Both tasks have been used frequently within the categorical perception framework.

In AX discrimination, two identical or different stimuli are presented within each trial, inviting “Same” or “Different” responses. An advantage of AX is that only two stimuli are presented, which means that the memory load is relatively low compared to discrimination tasks with more stimulus intervals (e.g. Wood, 1976). However, a disadvantage of AX is that listeners tend to favour one response (e.g. the response “Same” if the acoustic differences between stimuli are rather small) over the other, which leads to an unwanted response bias. Changes in response bias can be eliminated analytically by using signal-detection analysis of AX discrimination performance and are also diminished by using feedback about correct responses which reduces listeners’ uncertainty.

In AXB discrimination, listeners have to indicate whether the second stimulus is identical to the stimulus in the first or third interval. This task is a variant of ABX, but the same perceptual processes are expected to determine discrimination performance. According to Fujisaki & Kawashima (1970) the listener covertly classifies each stimulus in ABX, and if no decision is possible on the basis of phonemic information (i.e. if the first two signals in an ABX trial are classified as the same), the “timbre” of the stimuli serves as the basis for judgement. Thus, the auditory process is not consulted unless the phonemic process fails to yield a result, which is when A and B are classified as belonging to the same phoneme category.

In the present study, AXB is preferred over ABX, because previous studies have shown that listeners reduce uncertainty in an ABX task by comparing only B and X (e.g. Pastore, 1987). However, the same problem may occur in AXB. In Van Hessen & Schouten (1999) response times were measured that showed that, in AXB, listeners responded before the B stimulus had begun. If the B stimulus is ignored, the AXB-task is reduced to an AX-discrimination task. In the present experiment, no reaction times were measured, but we expect discrimination performance to be similar for AX and AXB if listeners decide to ignore the B stimulus in AXB discrimination.

As for the relationship between AX and AXB discrimination and classification, or the degree of categorical perception, we expect results to be similar to the findings with vowel stimuli in Schouten and Van Hessen (1992) and Van Hessen and Schouten (1999). This means that AX and AXB discrimination are expected to contain a peak that coincides with the peak in the classification data, but also that performance will be much better than predicted.

3.4.1 Method

The stimulus material was again the word vowel continuum and the 14 subjects from experiments I and II also participated in the present experiment.

3.4.1.1 Design

The experiment consisted of 4 tests: fixed and roving AX, and fixed and roving AXB. Each of the 7 stimulus pairs was presented 64 times.

In AX discrimination, stimuli were presented in four possible combinations: AA, AB, BA, and BB. Each combination was presented 16 times. Listeners had to indicate if the stimuli within a trial were identical or different. The inter-stimulus interval was 200 ms.

In AXB discrimination, stimuli were presented in the following four possible combinations: AAB, ABB, BAA, and BBA. Each combination was presented 16 times. The listeners had to indicate if the stimulus in the second interval was similar to the stimulus in the first or in the third interval. The inter-stimulus interval was 200 ms.

The classification results from experiment I were also used here, since the subjects were identical. Classification involved a forced choice between two alternatives, the vowels /u/ and /i/. Each stimulus (8) had to be identified 64 times in a random order.

3.4.1.2 Procedure

The stimuli were presented to the subjects over headphones in a sound-treated booth. In the discrimination tests, it was stressed that differences between the stimuli would be small, and in most cases could only be detected by listening carefully to all details of a stimulus. Each discrimination task was preceded by a training session, consisting of 56 trials. In the fixed discrimination context, there were 10 extra trials at the beginning of each block, which were considered practice and were not included in the data analysis.

The subjects responded by mouse clicking on one of two response fields on a computer screen. The labels of these response fields varied as a function of the discrimination task. In AX the response labels were “Same” and “Different” and in AXB “1” and “3”. After answering, visual feedback (250 ms) of the correct answer was given so that the subject was able to judge and possibly improve his or her performance.

In classification, one stimulus was played on each trial, and the subject had to identify it by mouse clicking on one of two response fields, “oe” or “ie” on a computer screen. The only training consisted of 16 trials. No feedback about correct responses was given.

3.4.2 Results AX and AXB

The AX and AXB discrimination and classification results are displayed in figure 3.11. The classification results are those from experiment I. The data in the figure represent the averages of the 14 subjects’ individual d' scores. The numbers (n) along the abscissa refer to stimulus pairs, consisting of stimuli (n) and (n+1). Stimuli in pair 1 resemble /u/ and stimuli in pair 7 sound like /i/. The calculation of the d' scores for discrimination was done according to the models for the separate discrimination tasks in Kaplan, Macmillan and Creelman (1978). The values of p(H) and p(FA) were forcibly limited to the 0.99 to 0.01 range, which meant that the maximum d'-value that could be obtained was 6.93 for AX discrimination, and it was 5.80 for AXB. For Classification it was 4.65.

As is illustrated in figure 3.11, AX and AXB discrimination show no close relationship with classification. Only the shape of roving AX seems to have a very rough similarity with the shape of the classification function, but AX discrimination is much better than predicted.

A two-way analysis of variance on the data revealed a significant main effect for the factor Task (5 levels) with F=12.60 and p<.001. The factor Stimulus (7 levels) had a small effect, F=2.59, p=.019. Post-hoc Tukey HSD on the factor Task showed that AX discrimination was significantly better than AXB discrimination, which again was better than classification. Post-hoc Tukey HSD on the factor Stimulus revealed that stimulus pair 3 differed significantly from the other pairs, which is probably due to the peak in classification performance.

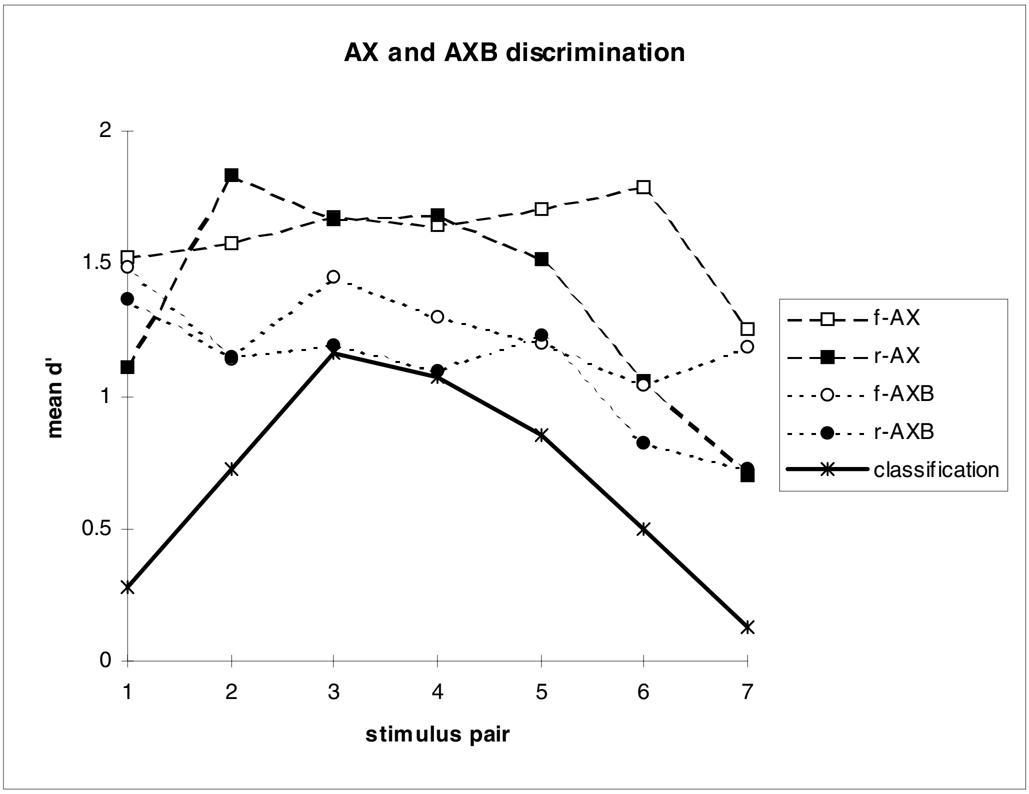


Figure 3.11. Classification and AX- and AXB discrimination results for the word vowels.

To test for the significance of the classification and discrimination peaks and valleys, one-way analyses of variance and post-hoc Tukey HSD tests were carried out on each separate task with Stimulus as independent variable. For classification there was a significant peak at stimulus pairs 3 and 4 (F=6.17, p<.001). There were no significant peaks or valleys in AX and AXB discrimination, i.e. not even in AX-roving, contrary to what might be suggested by the pattern in the figure.

If we compare AX with AXB discrimination, we see that performance differs between these two tasks. This implies that listeners have processed all three stimuli in AXB discrimination, instead of reducing AXB to AX or XB as was indicated by results of Pastore et al. (1976) and Van Hessen and Schouten (1999).

As in experiments I and II, we divided the results into quartiles on the basis of the 4I-oddity results. The previous results showed large differences among quartiles for the two 4-interval discrimination tasks, no difference for classification and a small difference for 2IFC discrimination.

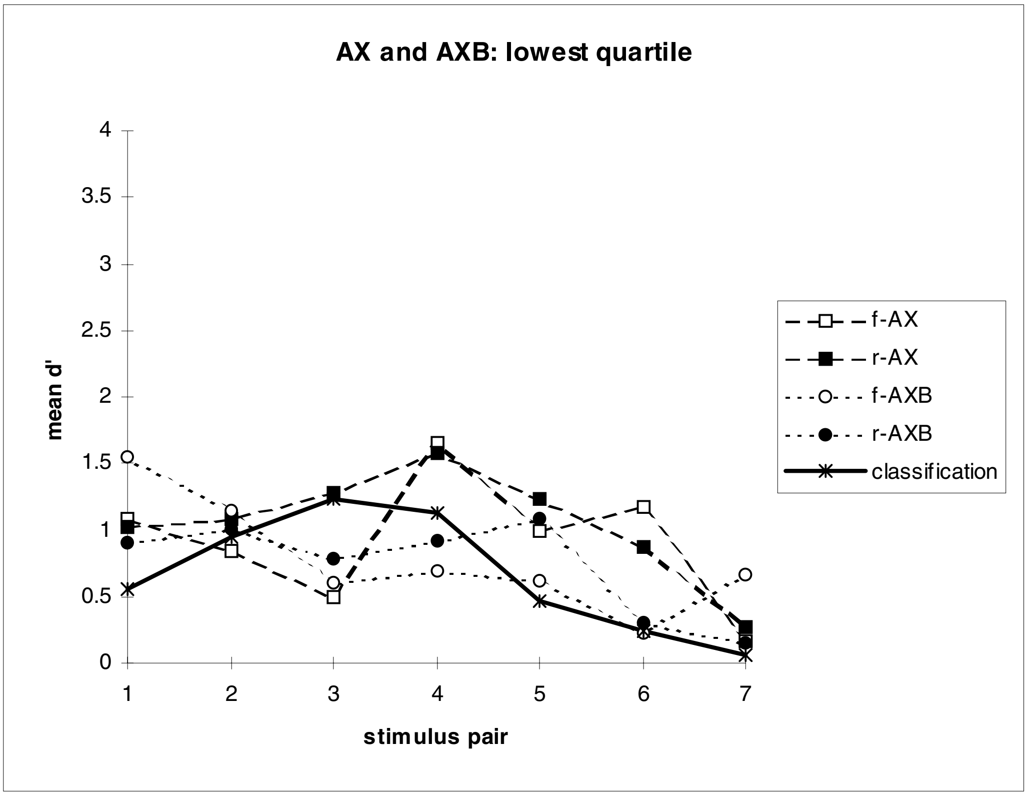


Figure 3.12 Vowel classification results and AX- and AXB discrimination of the lowest quartile (4 subjects).

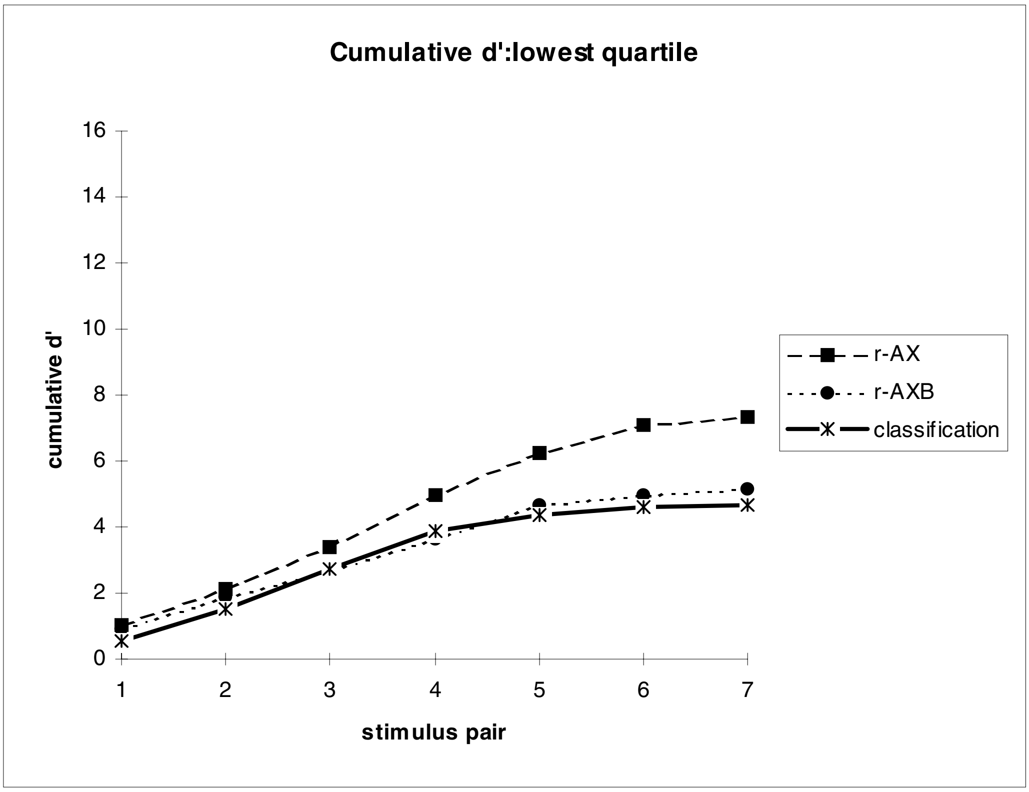


Figure 3.13 Cumulative d' results of the lowest quartile.

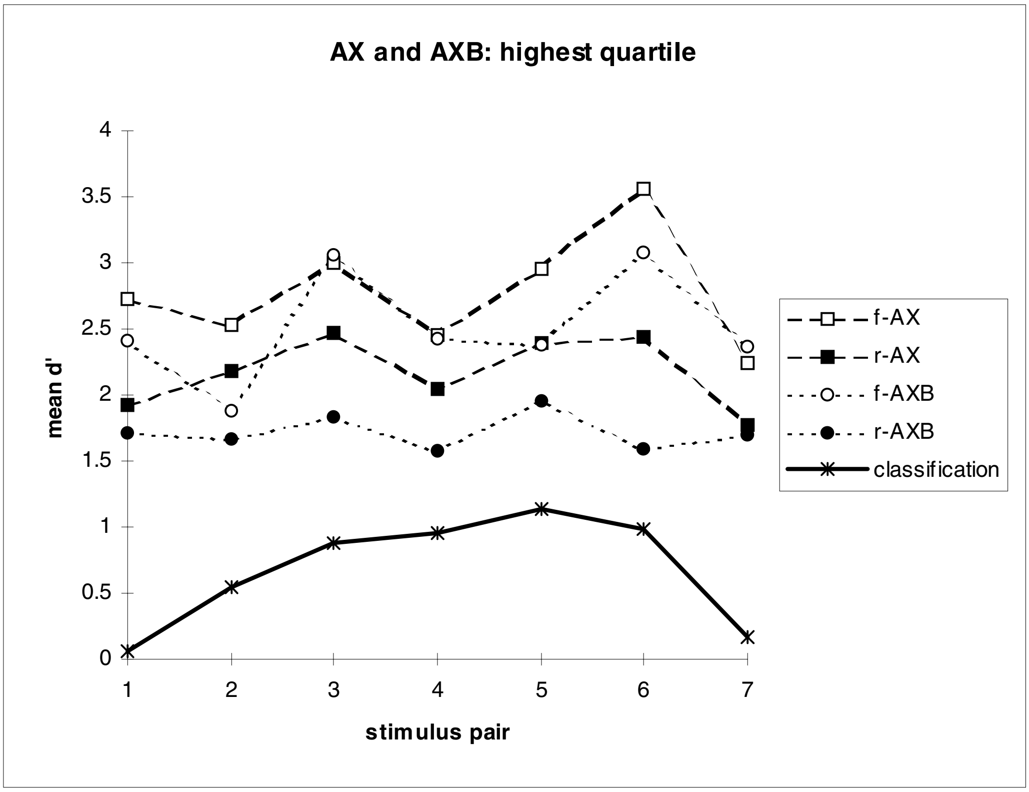


Figure 3.14 Vowel classification results and AX- and AXB discrimination of the highest quartile (4 subjects).

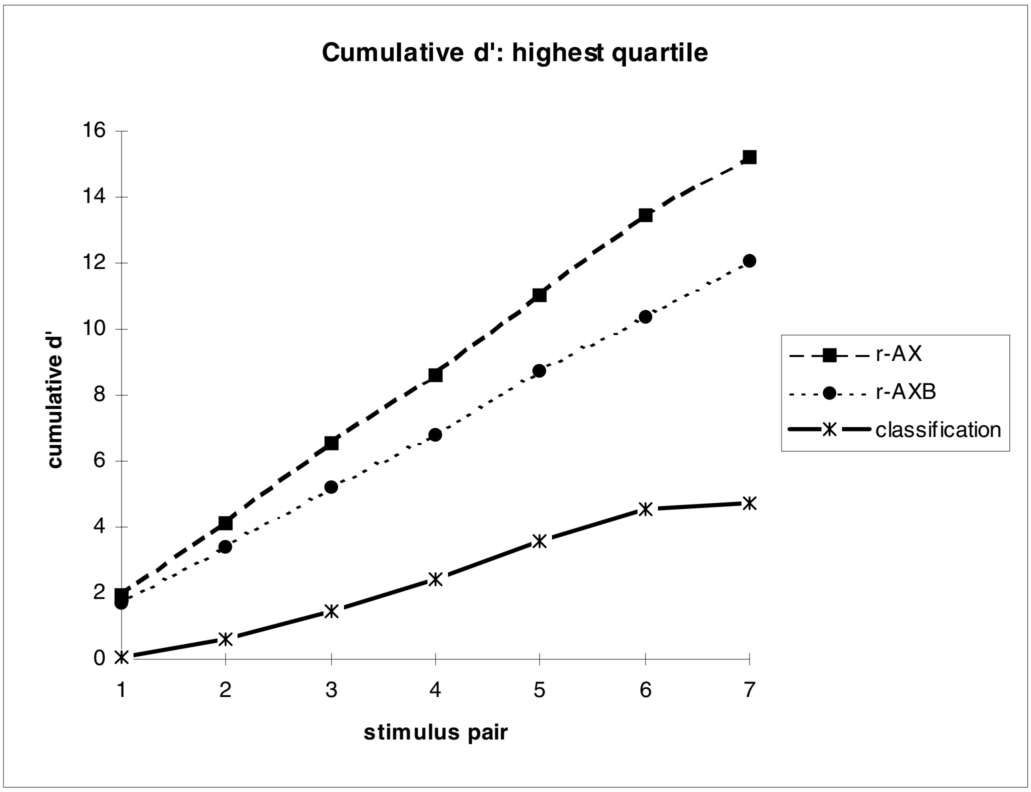


Figure 3.15 Cumulative d' results of the highest quartile.

In figures 3.12 to 3.15 the average and cumulative results of the lowest and highest quartiles are displayed. For AX and AXB it was predicted that there would be significant differences between the performance of listeners in the lowest and highest quartiles if they used mainly auditory-trace processes. Otherwise, if phonetic labelling processes had played a role in discrimination, it was expected that the lowest and highest quartile would not differ in their performance. A one-way analysis on the data with Quartiles (4 levels) revealed a significant main effect, F=46.83, p<.001. The quartiles differed significantly on AX and AXB discrimination, but not on classification performance. If we look at the results in the figures, figure 3.12 seems to show a peak in AX discrimination, which indicates that listeners have used a labelling strategy during this discrimination task. The individual subjects' results revealed that this was indeed the case for 9 of the 14 listeners and thus that there was again large between-subject variation, with some listeners showing a clear labelling effect and others not. Furthermore, 6 listeners showed peaked results in AXB.

In figures 3.13 and 3.15, with the cumulative results, it can be seen that for both quartiles AX performance was better than AXB performance, which is in agreement with the statistical analysis of the overall results. In figure 3.13, it seems as if the cumulative AXB performance for the lowest quartile almost completely coincides with cumulative classification. The resemblance of the two functions in the cumulative results is probably caused by the large scale of the vertical axis, which makes small differences less visible. In figure 3.12 it is evident that the shape of these two functions is much less related. Altogether, these findings indicate that in AX and AXB performance, listeners tend to use both auditory-trace coding and phonetic labelling processes.

Our AXB results show less phoneme labelling than in previous studies. In Van Hessen and Schouten (1999), AXB for vowels was fairly categorical. Differences between their study and the present one are a difference between the vowels, which were /i-e-a/ in Van Hessen and Schouten, plus the fact that Van Hessen and Schouten used trained listeners (phoneticians). MacKain, Best and Strange (1981) used the AXB paradigm to test discrimination between stimuli in a /r-l/ continuum. Their AXB results showed a discrimination peak that roughly coincided with the peak in classification and within-category discrimination was much better than predicted. However, there are two factors in their study that may have encouraged listeners to label the stimuli: a rather long ISI of 1000 ms and the lack of feedback about correct responses.

In summary, the findings from experiment III show that there is no strong relationship between AX or AXB discrimination performance and classification, and hence no categorical vowel perception. But this lack of a relationship does not mean that listeners relied only on auditory-trace processes as they did in 4I-oddity; they used a combination of phoneme labelling and auditory-trace coding. For a relatively large group of subjects, the shape of the AX discrimination function (and to a lesser extent for AXB discrimination) coincided with the shape of the classification function, but discrimination was also much better than predicted, especially for the within-category stimuli. Differences between listeners in performance on AXB and AX tasks are also reported by Repp (1981). In his study some subjects perceived /σ-Σ/ stimuli in a continuous manner, far better than predicted and without a discrimination peak, whereas others showed pronounced peaks at the category boundary and also better than predicted within category discrimination.

The results of this experiment support our basic conclusion that the degree of categorical perception varies as a function of the discrimination task. Moreover, these results substantiate the effect of the listener as a factor causing variance in the degree of categorical perception, which we also found in experiment II.

3.3 Discussion and conclusion

This chapter reports a series of experiments in which the effects of the discrimination task and of the inter-stimulus interval on the degree of categorical perception have been tested. Experiments I and III describe the relationship between classification performance and performance on five different paradigms for speech discrimination: 2IFC, AX, AXB, 4I-oddity and 4IAX discrimination. The results show that there are major differences between discrimination results as a function of task, which means that the degree of categorical perception varies due to task differences.

Of all the five discriminations tasks, 2IFC-discrimination demonstrated the closest relationship with classification performance, which proves that phonetic labelling plays a major role during this task. However, 4I-oddity discrimination with the same stimuli showed no relationship with classification at all. This difference between 2IFC and 4I-oddity lends strong support to our initial interpretation of the 4I-oddity results: performance is based on an auditory-trace strategy only. In general, the results support our predictions about the perceptual processes during discrimination, although there appeared to be a rather large between-subject variation in AX, AXB and 4IAX: some listeners used a labelling strategy which was revealed by a peak in the discrimination data, whereas others did not. AXB turned out to be more like an economic variant of 4IAX instead of being a variant of ABX (see also Repp, 1984). This may be caused by the short interval (200 ms) between the three stimuli in our AXB task. In previous studies with ABX, the temporal interval between X and B was often longer than between A and X, which may induce phonetic labelling (e.g. Pisoni & Lazarus, 1974).

A comparison of more than two paradigms for speech discrimination in a single study has not been done before. Creelman and Macmillan (1979) made an extensive comparison of different paradigms for nonspeech discrimination. They found other levels of sensitivity for the separate tasks than we did, which suggests that speech-nonspeech stimulus factors interact with task factors. Creelman and Macmillan (1979) also found that fixed discrimination was always superior to roving discrimination, which is confirmed by our study (see also Schouten & Van Hessen, 1992).

The effect of the duration of the temporal interval between stimuli was tested in Experiment II. The longer inter-stimulus interval facilitated discrimination, but the effect on the degree of categorical perception was relatively small. 4I-oddity discrimination improved as an effect of the longer ISI, but still showed no relationship with the classification results. Apparently, 4I-oddity is highly sensitive to acoustic differences between the stimuli, which does not change when the ISI duration changes from 200 to 500 ms. This may sound contradictory to findings in previous studies in which long ISI’s tend to encourage labelling as a consequence of the decay of auditory memory. However, our findings suggest that this effect may only occur at even longer ISI’s, like 1 sec, and also that this effect may be task-specific. In our case, more labelling was only found for 4IAX, in which more listeners showed a peak in performance with the longer interval than with the shorter one. In this task an ISI of 500 ms was used, but the interval between pairs was 1 sec. 4IAX results were still better than predicted, which indicates that in this task the auditory trace was also used, despite the longer temporal interval between stimuli. These results appear consistent with the vowel data of Crowder (1982) and Repp, Healy and Crowder (1979), who showed that, at a long interval observed 4IAX discrimination scores were higher than predicted.

In conclusion, the findings indicate that if a discrimination task is used in which listeners are stimulated to compare the auditory traces of the stimuli, discrimination performance will show no relation with classification performance. If a discrimination task is used during which listeners use mainly phoneme labelling to discriminate between stimuli, discrimination performance will show a close relation to classification performance. On the basis of these results it has to be concluded that the relation between classification and discrimination results simply reflects the degree to which listeners use the same perceptual processes during these tasks. Discrimination of two speech sounds on a stimulus continuum has nothing to do with classification of speech sounds and hence with categorical speech perception. However, up to now the speech sounds under investigation in this study have been vowels, of which it is said that discrimination performance varies more than for consonants. Therefore we compared vowel discrimination with discrimination of the most prototypical categorical perception stimuli: stop consonants. These experiments are described in the next chapter.

Chapter 4

Categorical perception of stop consonants and vowels

4.1 Introduction

In the preceding chapters we have shown that the degree of categorical perception of vowels varies as a function of the discrimination task. Perception of vowels can be very categorical, but it can also be completely non-categorical. We conclude from these results that an unvarying relationship between discrimination and classification of speech sounds does not exist.

However, supporters of the categorical perception model may find the observation that vowel discrimination varies as a function of the task not very convincing as an argument for rejecting their model. They may argue that it is not unusual for the degree of categorical perception to vary with vowels (e.g. Pisoni, 1975). Furthermore, they may claim that only consonant stimuli are perceived categorically, not vowels. Although the results with 2IFC discrimination (chapter 3) have clearly shown that there can be a relationship between observed and predicted vowel discrimination, it would be useful to substantiate our conclusion with data from a different phoneme contrast.

Categorical perception of stop consonants is often cited as a robust phenomenon, which has been found in many settings (see several reviews in Harnad, 1987). Even Massaro (1994), who is one of the prominent sceptics with respect to categorical perception, reluctantly admits that “there may be a few phoneme contrasts that qualify for a weak form of categorical perception.” The few phoneme contrasts Massaro (1994) refers to are stop-consonant ones.

To explain the high degree of categorical perception for stop consonants, Liberman, Mattingly and Turvey (1972) suggested that the listener has no auditory image of the signal available, but only the output of a specialised processor that has stripped the signal of all normal sensory information and represents each phonetic segment categorically. This claim has motivated researchers to prove the opposite, which is that listeners do hear differences between within-category stop consonant stimuli.

Firstly, reaction-time measurements have been used to show that listeners are sensitive to differences between within-stop-category stimuli (Massaro, 1987; Pisoni & Tash, 1974; Repp, 1981). According to the categorical perception model, listeners cannot discriminate differences within a category and therefore there is no basis for differences in reaction times to different stimuli. However, in Pisoni and Tash (1974) it was shown that the time it took subjects to respond “Same” to different within-category pairs of stimuli was significantly longer than the time to respond “Same” to pairs in which the stimuli were actually identical. This finding suggests that the auditory system did detect some differences between sounds. In Massaro (1987), reaction times were measured for classification of a stop consonant place continuum. Massaro’s results revealed an increase in reaction times near the category boundary, which implies that stimuli within a stop-consonant category varied in ambiguity or the degree to which they represented the category. However, Samuel (1979) and Pisoni and Tash (1974) have failed to find different labelling reaction times for stop-consonant stimuli that were strictly within the same category.

Secondly, rating experiments have shown that listeners give different ratings to within-category stimuli. Apparently, they feel that these stimuli differ in the degree to which they represent the phoneme category (Massaro & Cohen, 1983).

Thirdly, it has been demonstrated that, although certain speech continua were perceived categorically by naive subjects, after discrimination training with the given continuum listeners were able to resolve very fine distinctions between within-category pairs (Carney, Widin & Viemeister, 1977; Pisoni et al., 1982; Samuel, 1977; Werker & Logan, 1985).

The rating, reaction-time and training studies demonstrate that, to some extent, listeners have access to auditory differences between within-stop-category stimuli. As for the training studies, however, there have also been several experiments in which extra training failed to improve within-stop-category discrimination (see review Strange & Jenkins, 1978). Furthermore, the extensive training and the use of highly practiced subjects is an important drawback of the training studies, since it does not show what listeners typically do in everyday speech perception. For instance, in Carney et al. (1977) three subjects, all with extensive experience in previous psychophysical experiments, were tested in more than a dozen sessions distributed over several weeks. It would be interesting to see if within-category discrimination could be improved without using highly experienced listeners and without numerous training sessions.

Only Pisoni and Lazarus (1974) have tried to demonstrate that within-category perception of stops can vary with tests that are used in the traditional categorical perception framework. They tested the perception of stimuli within a voice-onset continuum by comparing a classification task with ABX and 4IAX discrimination tasks. The results showed that both ABX and 4IAX discrimination were consistently better than predicted discrimination, although the shape of the observed and predicted functions was fairly similar. The level of within-category discrimination was slightly higher under the 4IAX test than under the ABX test, but only for two-step comparisons and not for one-step comparisons. At the peaks in the functions, performance in both discrimination tasks was comparable. According to Pisoni and Lazarus (1974), the two-step 4IAX results imply that listeners can discriminate between within-category stimuli in a non-phonetic mode. However, this study also included a condition in which the subjects were given extra training before performing the discrimination tests. This extra experience with the stimuli appears to have been the critical factor in the results, for Pisoni and Glanzman (1974) failed to find any difference between stop consonant discrimination tested with the ABX or 4IAX tasks when no pretraining was provided.

The aim of the present study was, firstly, to see if degree of categorical perception of stops could be varied in a design similar to the one used by Pisoni and colleagues and, secondly, to compare stop-consonant perception with vowel perception. We compared classification performance with discrimination performance in two different discrimination tasks: 2IFC and 4IAX discrimination. By using 2IFC instead of ABX, we expected to find more differentiation between the discrimination results than Pisoni did. We used 4IAX instead of 4I-oddity since both tasks seem to be sensitive to the acoustic differences between stimuli, but 4IAX has the advantage of having a standard signal detection model with which d' can be calculated.

Our previous experiments have shown that, in 2IFC discrimination, listeners are strongly encouraged to use their internal phoneme representations to make a decision about the order of the stimuli, whereas 4IAX discrimination is mainly based on an auditory trace strategy (e.g. Pisoni, 1975; Schouten & Van Hessen, 1992). Thus, for 2IFC, observed discrimination is expected to be highly predictable from classification. For the 4IAX results we predict that observed discrimination will be less clearly related to the classification data.

Furthermore, we hypothesise not only that stop consonant discrimination will vary as a function of the discrimination task, just as was the case for vowels, but also that the stops will be perceived more categorically than the vowels, as has been shown by numerous previous studies (e.g. Fry et al., 1962; Pisoni, 1973; Repp, 1981; Stevens et al., 1969). Thus, although the relationship between observed and predicted discrimination may be different for the two discrimination tasks, there will always be a stronger relationship between the results for stop consonants than for vowels.

4.2 Method

4.2.1 Stimulus material

The speech material in this experiment consisted of seven stimuli in a stop consonant /pAk/-/tAk/ continuum and eight stimuli in a vowel /pup/-/pip/ continuum. The stop-consonant and vowel stimuli sounded like completely natural speech signals.

The /pAk/ - /tAk/ stop consonant continuum was similar to the one used in the experiments described by Van Hessen and Schouten (1999). The generation of the stop-consonant continuum was done with linear interpolation between the relative amplitudes of the spectral envelopes of /pAk/ and /tAk/ according to the method described in chapter 2. The original words were produced by a male native speaker of Dutch. Source spectra and filter of the original words were determined and separated frame-by-frame. The frames were 25.6 ms long and the frame shift was 6.4 ms, resulting in a great deal of frame-to frame overlap. Source spectra and spectral envelopes were described by means of the phases and amplitudes of 100 spectral components. The interpolation between the original stop consonants consisted of six steps and consequently resulted in seven stimuli.

The vowel stimuli were the ‘text vowels’ used in the experiments of chapter 2 (see section 2.2.1) and consisted of a 7-step continuum between /pup/ and /pip/. The original words had been read aloud in a written text at a fast speech rate by a male native speaker of Dutch.

4.2.2 Subjects

Twelve listeners took part in the experiments. In the vowel experiment, one subject's data were ignored because his responses turned out to consist largely of missing values, probably due to computer failure. Thus eventually there were 12 listeners for the stop consonant tests, but only 11 for the vowel tests. The subjects were all native speakers of Dutch who were paid a basic rate plus a reward depending on the number of correct responses they gave. The listeners in the current experiments were different from those in the previous experiments, since most subjects of the initial subject pool were no longer available by the time this test took place.

4.2.3 Design

All subjects took part in three different tasks on each of the two stimulus continua: fixed and roving 2IFC discrimination, fixed and roving 4IAX discrimination and classification. In one series of tests, the stop continuum between /pAk/ and /tAk/ was presented and in another series of tests the stimulus material consisted of the vowel continuum between /pup/ and /pip/. The order of presentation of the continua and the tests was balanced over subjects.

In 2IFC discrimination all trials consisted of two different stimuli, either AB or BA. The subjects responded by indicating the order in which the stimuli were presented. The inter-stimulus interval was 300 ms and response time was 1200 ms. In the fixed context each pair was presented repeatedly in one block of trials. For the vowel stimuli there were 7 blocks, and for the stops there were 6 blocks, one block for each pair. Each block contained 32 trials, 16 for each of the two possible combinations. In the roving discrimination experiment, the A and B stimuli to be discriminated were chosen randomly from the total range of stimuli and thus varied from trial to trial. In the roving 2IFC test for vowels, 7 (stimulus pairs) x 32 trials were presented, and in roving 2IFC for stop consonants, 6 (stimulus pairs) x 32 trials were presented.

In 4IAX discrimination, each trial consisted of two stimulus pairs, one of two identical stimuli and the other of two different stimuli; AB-AA, BA-AA, AA-BA, AA-AB, BA-BB, AB-BB, BB-AB, BB-BA. These 8 combinations were presented 4 times in random order. In the 4IAX test, subjects reported which pair of stimuli was the same, the first pair or the second pair. ISI was 300 ms, the interval between pairs was 600 ms and response time was 1200 ms. In the fixed context, all 32 trials of one, randomly chosen, stimulus pair were given before the next stimulus pair was presented. For the vowel stimuli there were 7 fixed blocks, and for the stops there were 6 blocks. In the roving context stimulus presentation was random.

Classification involved a forced choice between two alternatives, the vowels /u/ and /i/ or the stop consonants /p/ and /t/. There were separate tests for each of these two continua. All stimuli were presented to subjects 32 times in completely random order. Response time was 1200 ms.

4.2.4 Procedure

The stimuli were presented to the subjects over headphones in a sound-treated booth. In the discrimination tests, it was stressed that differences between the stimuli would be small, and in most cases could only be detected by listening carefully to all details of a stimulus. Subjects were paid two cents for each correct response and had the same amount deducted for each mistake (this meant that random respondents were paid exactly the basic hourly rate).

The subjects responded by pressing the mouse button on one of two response fields on a computer screen. In 2IFC, the response fields were labelled "ie-oe" or "oe-ie" (/ɩ-ʊ/ or /ʊ-ɩ/) for the experiment with the vowel stimuli. In the stop consonant continuum the 2IFC labels referred to the stop contrast. In 4IAX, the response labels were "pair 1" or "pair 2". During discrimination listeners received visual feedback (500 ms) of the correct answer after each response. Discrimination training consisted of 56 trials for the stop consonant continuum and of 64 trials for the vowel continuum. Training was intended to familiarise subjects with their task.

In classification, one stimulus was played on each trial, and the subject had to identify it by mouse-clicking on one of two response fields (phoneme labels) on a computer screen. The only classification training consisted of 14 (stops) or 16 (vowels) trials. No feedback about correct responses was given.

4.3 Results

The results of the experiments with stop consonant and vowels are presented in sections 4.3.1 and 4.3.2. The figures display the listeners' average observed and predicted discrimination performance. The results of the classification tests are presented as predicted discrimination scores. The 2IFC and 4IAX discrimination d' scores were calculated by subtracting z(FA) from z(H) and then transforming these scores into 'true' d'-values (Kaplan, Macmillan & Creelman, 1978; Macmillan & Creelman, 1991; Macmillan, Kaplan & Creelman, 1977). The values of p(H) and p(FA) were limited to the 0.99 to 0.01 range, which meant that the maximum d'-value that could be obtained was 3.29 for 2IFC. For 4IAX it was 5.14, and the maximum d'-value for classification was 4.65.

4.3.1 Stop-consonant results

The mean stop-consonant results are displayed in figure 4.1. The stop-consonant stimuli in pair 1 resemble /p/ and the stimuli in pair 6 sound like /t/. The bold line represents predicted discrimination/classification, the dotted lines represent the 2IFC discrimination results and the thin solid lines are the 4IAX results.

Figure 4.1 leads to the following observations: (1) there is no strict relationship between the results of either of the two discrimination tasks and the classification results; (2) fixed 2IFC contains a small peak but is very low, even lower than roving 2IFC. The roving 2IFC discrimination function appears to contain a discrimination peak which roughly coincides with the classification peak, and within-category discrimination is similar to classification performance; (3) the shapes of the 4IAX discrimination functions are unrelated to the classification function and within-category discrimination is much better than predicted. Fixed 4IAX seems to have two small discrimination peaks, the second of which coincides with the peak in 2IFC discrimination. However, it is very likely that these fixed 4IAX peaks are not significant.

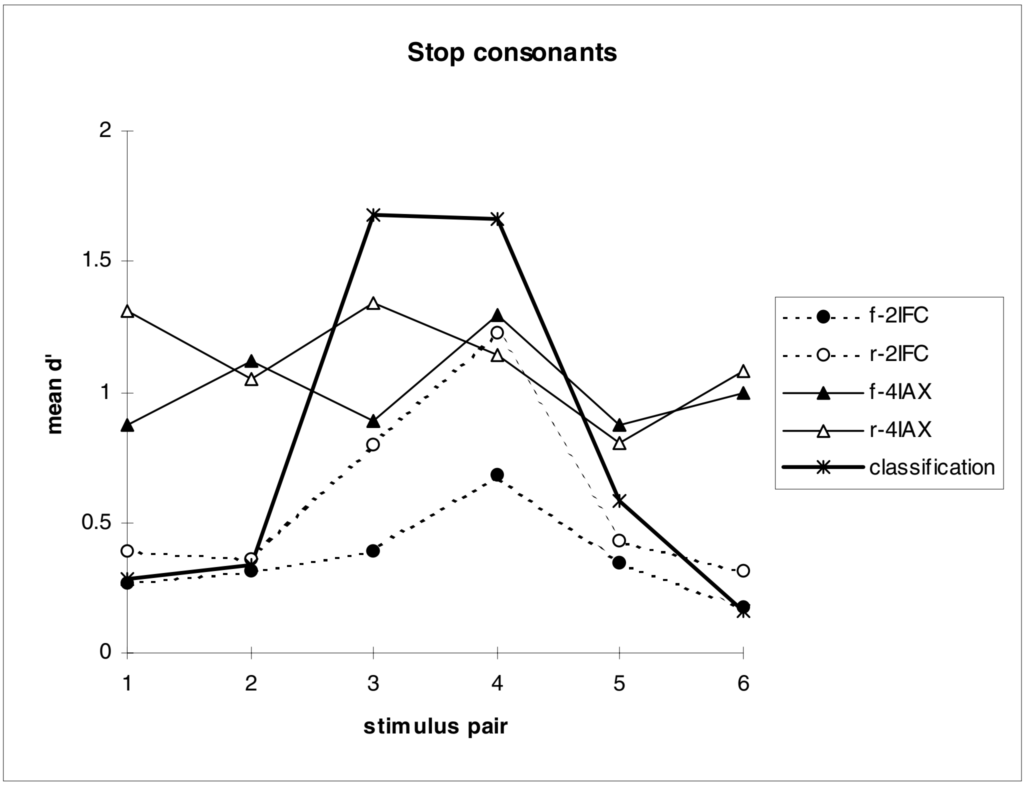


Figure 4.1. Classification and 2IFC and 4IAX discrimination results for the stop consonants.

Statistical analysis

A two-way analysis of variance was performed on the observed and predicted discrimination data. The two main effects under consideration were: Task (5 levels) and Stimulus (7 levels). Cell variance was over 12 subjects. The results of the analysis showed a significant difference between the tasks, F=32.87, p<.001. Discrimination performance was always better with the 4IAX tests than with the 2IFC tests or than predicted by classification. The main effect for Stimulus was also significant, F=20.97, p<.001. Furthermore, the Task by Stimulus interaction was significant, F=5.05, p<.001, indicating that the discrimination behaviour of the listeners varied as a function of the task at different points along the stimulus continuum.

To test for the significance of classification and discrimination peaks, separate one-way analyses and post-hoc Tukey HSD tests were conducted on the data of each task, with Stimulus as independent variable. There were no significant peaks in the fixed and roving 4IAX results. However, the analyses did reveal a significant peak at pair 4 for fixed and roving 2IFC discrimination (F=3.24, p=.011 and F=9.43, p<.001) and also at pairs 3 and 4 for classification (F=29.22, p<.001). The 2IFC results imply that listeners tried to use phoneme labels to discriminate the stimuli, but the results in figure 4.1 show that they were not completely successful in doing so, because at the discrimination peak 2IFC discrimination is lower than predicted by the labelling data. At pair 3, both fixed and roving 2IFC are significantly lower than predicted discrimination, F=14.27, p<.001. At pair 4 only fixed 2IFC is lower, F=3.319, p=.017. The data show that fixed 2IFC is worse than roving 2IFC discrimination. However, listeners performance is usually better in the fixed than in the roving context. We will return to this issue in the discussion of the stop-consonant results, section 4.3.1.2.

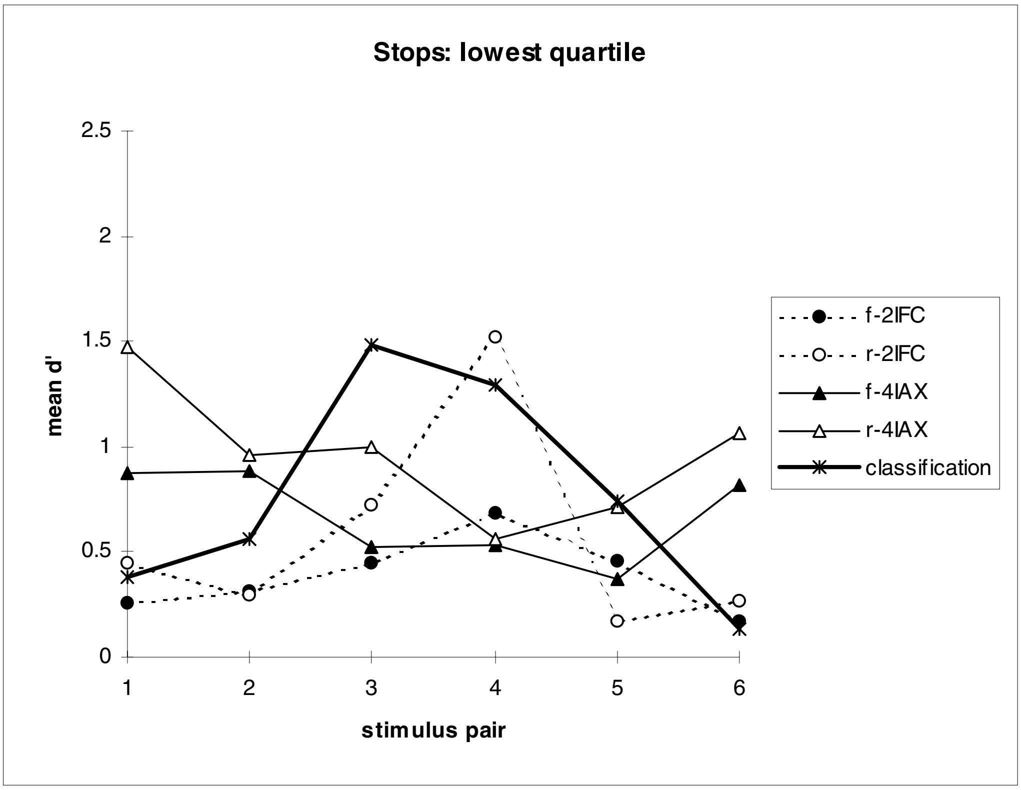


Figure 4.2. Stop-consonant classification and 2IFC and 4IAX discrimination results of the lowest quartile (3 subjects).

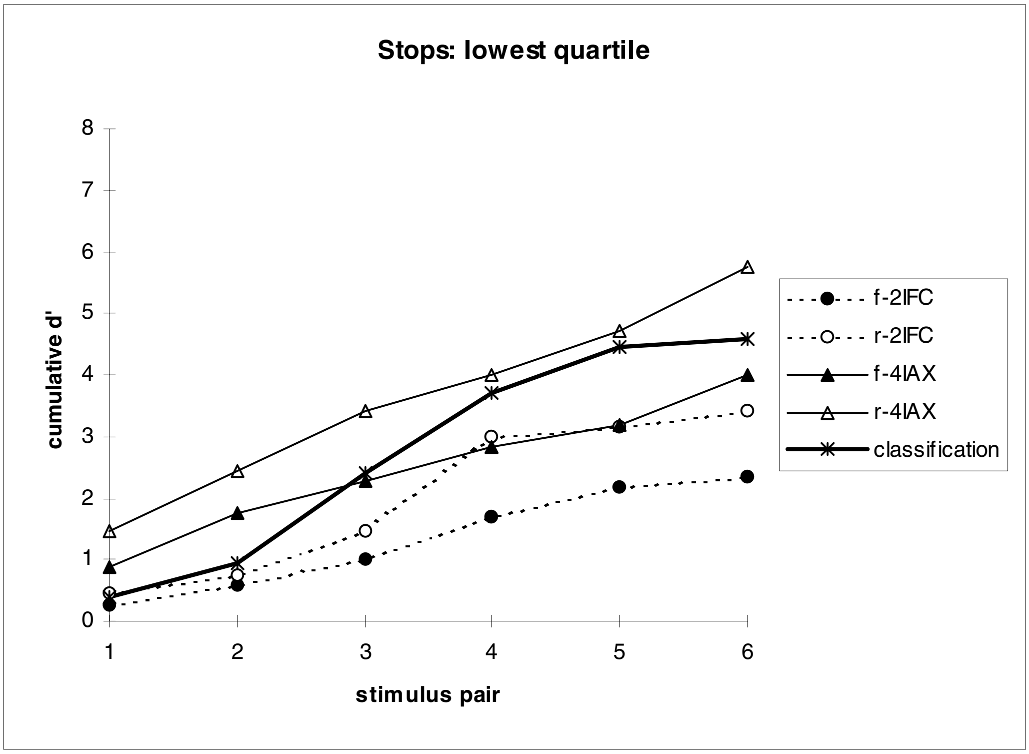


Figure 4.3. Cumulative stop-consonant classification and discrimination of the lowest quartile.

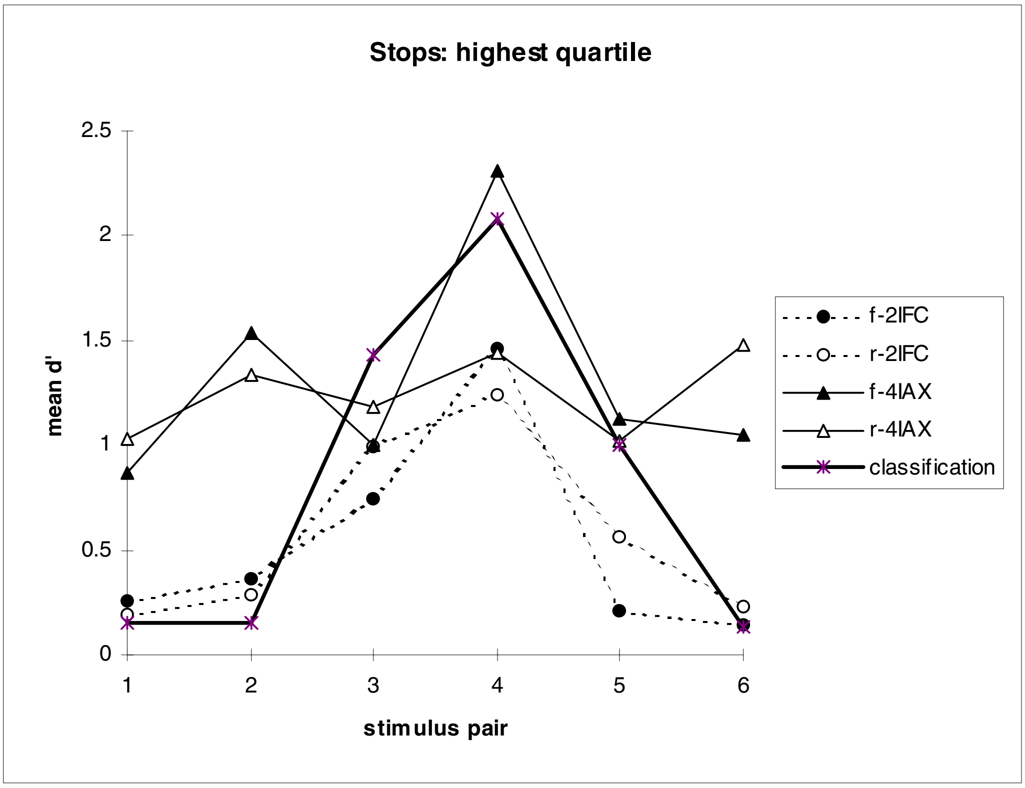


Figure 4.4. Stop-consonant classification and 2IFC and 4IAX discrimination results of the highest quartile (3 subjects).

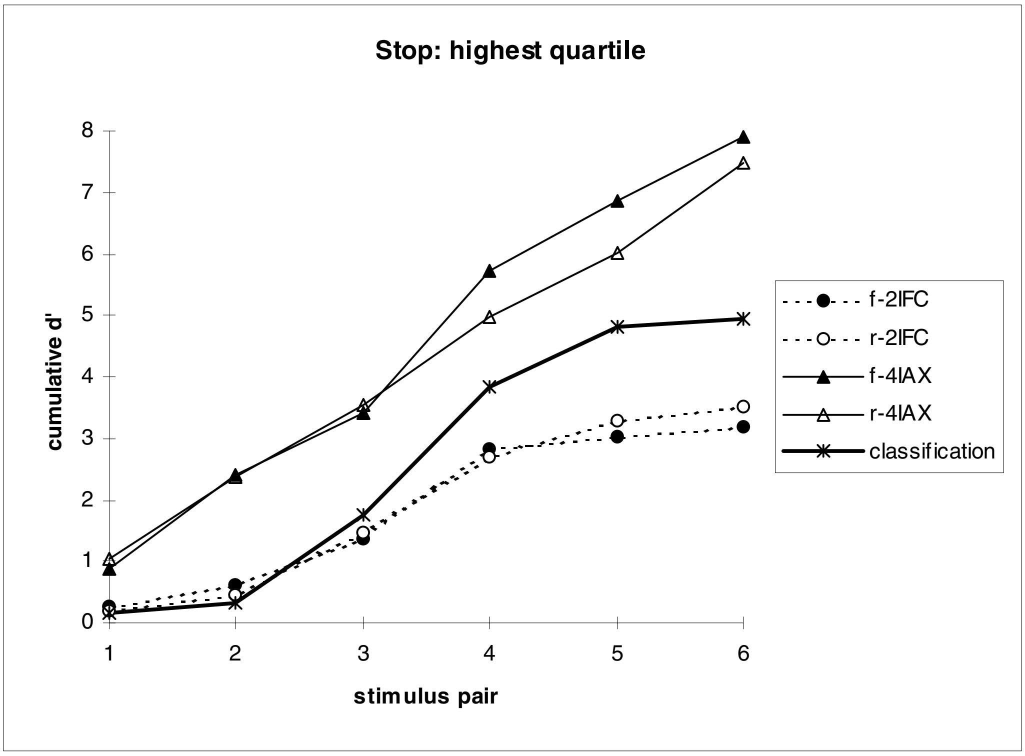


Figure 4.5. Cumulative stop-consonant classification and discrimination of the highest quartile.

Post hoc Tukey-HSD analysis of Task at each stimulus pair showed that fixed and roving 4IAX were significantly higher at all pairs except at pair 3, at which roving 4IAX and Classification were highest, and at pair 4, at which Classification was significantly higher than all other tasks. This indicates that within-phonetic-category comparisons benefited more from the 4IAX procedure than across-category comparisons, relative to the predicted discrimination performance.

4.3.1.1 Differences among subjects

In the stop-consonant experiment 54% of the total variance was explained by cell variance. As in the previous experiments, we took a closer look at listeners' performance by dividing the pool of subjects into quartiles. This time the criterion was based on fixed and roving 4IAX discrimination. Figures 4.2. and 4.4 display the mean results of the three subjects in the lowest quartile and the highest quartile, respectively. To get a better view of the differences between performance of the quartiles, the cumulative d' values are also presented (figures 4.3 and 4.5). The cumulative d' scores, which were calculated simply by adding up the d' values from left to right, give an indication of the overall level of performance as a function of task.

It has to be noted that there were only 3 subjects in each quartile, which is also the reason why we ran only one statistical analysis on this subset of data. The effect of Quartiles (4 levels) was tested on the data per task and was only significant for fixed 4IAX ($F=4.62$, $p<.005$). This is confirmed by the cumulative functions in figures 4.3 and 4.5 in which fixed 4IAX is generally worse than predicted in the lowest quartile, but much better than predicted in the highest quartile.

Figure 4.2 shows the averaged results of the three subjects in the lowest quartile. However, the results of each of the three subjects within this quartile differed considerably from this average pattern. One of the subjects showed a close relationship between fixed and roving 2IFC and classification. The second subject had results that are similar to the average pattern: roving 2IFC was very similar to classification. The discrimination results of the third subject were low across the board.

From figure 4.4, which displays the results of the highest quartile, a very different pattern emerges. However, the separate results of each of the three subjects within this quartile were also very different from the average pattern. One subject showed a much better than predicted peak in the fixed 4IAX results at pair 4, but had low overall results for the other three discrimination tests. The second subject showed a flat shape for 4IAX discrimination and a close relationship between both fixed and roving 2IFC and classification. The third subject had two marked peaks in fixed 4IAX (the first of which was at pair 2) and a peak in roving 4IAX, which roughly coincided with the classification peak. This subject's 2IFC was low.

It seems better, therefore, to focus on the differences and similarities between the lowest and highest quartiles, than to try to interpret the peaks and valleys in the separate figures. There was only a difference between quartiles in tasks in which the listeners were expected to be encouraged to use auditory trace coding. Performance in phoneme labelling tasks, i.e. classification and 2IFC discrimination, was not significantly different.

4.3.1.2 Discussion stop-consonant results

The question we addressed in this experiment was if the degree of categorical perception for stop consonants would vary as an effect of the discrimination task. The results show that this is indeed the case: the degree of categorical perception for stops varies as a function of 2IFC or 4IAX discrimination. However, there is no strict relationship between classification and either 2IFC or 4IAX discrimination. Nevertheless, it is evident that 2IFC discrimination is more closely related to classification than is 4IAX discrimination. Thus, as we expected, responses are more categorical for 2IFC, since this task presumably encourages listeners to use a phoneme labelling strategy during discrimination. On the other hand, perception is less categorical for 4IAX discrimination, because this task is more sensitive to the acoustic differences between the stimuli.

However, this interpretation of the 2IFC results only accounts for the roving data. In the fixed 2IFC results we see d' values that are worse than the values for roving 2IFC. Usually, fixed context results are better than roving context results, regardless of the discrimination task (e.g. Schouten & Van Hessen, 1992; Van Hessen & Schouten, 1999). In fixed discrimination, listeners hear the same stimuli over and over again and thus hear only a small subset from the total range of stimuli in the continuum (see also chapter 2, section 2.1.5). It was expected that, as a consequence of the small range, stimulus uncertainty would be lower in fixed than in roving discrimination. This leads to a higher degree of labelling in the fixed condition. According to Durlach and Braida (1961) listeners will construct a temporary internal representation of the range of stimuli presented in a particular experiment, which in the case of speech sounds may coincide with phoneme categories in long-term memory. If the temporary labels represent phoneme categories, the results will represent the effect of phoneme labelling processes.

The present fixed 2IFC results are not in accordance with our expectations. It appears that listeners were unable to use any labelling at all. We can only speculate about the causes of this unusual behaviour, particularly since a labelling strategy *was* followed in roving 2IFC discrimination. Apparently, applying the two phoneme labels was easier when the full stimulus range was available than when only two very similar stimuli were presented from trial to trial so that listeners did not regularly hear clear exemplars from both categories.

However, it has to be noted that there was a large variation within the fixed 2IFC results due to individual differences. For 6 of the 12 subjects, fixed 2IFC was worse than roving 2IFC. For 4 subjects, fixed and roving 2IFC performance was similar and two subjects had better fixed than roving 2IFC. We have to conclude, therefore, that the use of labelling is not only dependent on the discrimination task itself, but also on the stimulus context and on the listener.

One could also argue that the interpretation of roving 2IFC is incorrect, because the scores are too low at the phoneme boundary (worse than predicted). A hypothetical explanation is given by Schouten and Van Hessen (1992), who found similar results with almost identical stimuli. The low 2IFC results motivated them to propose a new model for the calculation of d' for 2IFC discrimination. If we had used the model by Schouten and Van Hessen (1992), the present 2IFC results would have been $\sqrt{2}$ times as high, and this would have resulted in a very close relationship between observed and predicted discrimination and a high degree of categorical perception. In fact, the results would have been identical to those found for stop consonants by Schouten and Van Hessen (1992). However, we used the standard model by Macmillan et al. (1977). We will elaborate on the two models to explain their differences and our choice of the standard model.

In 2IFC discrimination, two different stimuli are presented and the task of the listener is to determine the order of the stimuli (AB or BA). According to Macmillan et al. (1977), the optimal strategy for the listener is to subtract the first stimulus from the second. Each of the two stimuli has, by definition, a normal distribution with a variance of 1. The subtraction of the two stimuli results in two new normal distributions with means that are equal to the differences (A-B and B-A) of the two original means, and variances that are equal to the sums of the two original variances. On the basis of the two response categories (AB and BA), the listener has to create an internal decision criterion. If the difference between the auditory traces surpasses this criterion, the listener decides to choose the order BA, if not, the order is AB.

According to Schouten and Van Hessen (1992), 2IFC discrimination of speech stimuli involves different decision processes than those proposed in the standard model, which is based on 2IFC discrimination of meaningless psychoacoustic stimuli like pure tones. In their view, a listener cannot decide about the order of speech stimuli, without assigning a ‘quality code’ to the stimuli first. This quality code is necessary, because differences between speech sounds are, unlike differences between pitch, duration and intensity, dependent on the phoneme categories they belong to. The quality code contains the phoneme label of the stimulus, and an estimate of the distance between this stimulus and its phoneme category. If the second stimulus receives the same phoneme label as the first one, a second quality code is created. After this coding, the order of the stimuli is determined in the way described for the standard model. This means that an extra step is added to the listeners’ decision strategy in the new model. As a consequence of this extra step, the comparison of stimulus and phoneme category, an extra source of variance is introduced. The variance of the phoneme category is estimated to be identical to the variance of the stimulus. As a result, a different calculation of d' for 2IFC discrimination is proposed, resulting in a ‘true’ d' that is √2 times as high as d' calculated according to the standard model. However, an important factor of dispute is the estimated variance of the phoneme category. If phoneme category representations in long term memory are perfect, variance equals 0. Because it is likely that the representations are less than perfect, the variance is expected to be higher than 0. However, it is unclear how much higher it really is. Assuming that the variance is 1 is a rather arbitrary choice; this is the reason why we preferred the standard model instead of risking that, in absolute terms, our 2IFC values would be too high.

The fact remains that, calculated either according to the standard or to the new model, our stop-consonant 2IFC results are almost identical to those found by Schouten and Van Hessen (1992).

As for the fixed and roving 4IAX data, there appeared to be large between-subject variation around the phoneme boundary location. In some of the results, fixed and roving 4IAX contained a marked discrimination peak, which coincided with the peak in classification. This peak was present in the fixed condition for 4 of the 12 subjects and in the roving condition for 2 of the 12 subjects. Overall, our 4IAX results are less categorical than the results of Pisoni and Lazarus (1974), which may be due to differences between the stimulus material. In their experiment, synthetic /bA/ and /pA/ syllables were used, which varied only in voice-onset-time. In the present study, a stimulus continuum spanning a stop place contrast was used in the meaningful words /pAk/ and /tAk/. The continuum was generated by interpolating between the spectral envelopes of the original words instead of interpolating between one specific parameter like VOT or onset formant frequency.

The next question of interest was whether these stop-consonant results would be different from vowel results in degree of categorical perception. We will return to this question in sections 4.3.2.2 and 4.4.

4.3.2 Vowel results

The discrimination and classification results for the vowel stimuli are displayed in figure 4.4. Vowel stimuli in pair 1 resemble /u/ and stimuli in pair 7 sound like /i/.

The general vowel patterns are very similar to those in the stop-consonant results. There is no strict relationship between observed and predicted discrimination performance for either of the two discrimination tasks. The fixed and roving 2IFC discrimination functions contain a discrimination peak, which roughly coincides with the classification peak, although this peak is very small in the fixed 2IFC function. Observed fixed 4IAX discrimination contains two peaks, the second of which seems to be related to the classification peak and to the roving 2IFC discrimination peak. Within-category 4IAX discrimination is much higher than predicted.

A minor difference between the stop and vowel data is that within-category 2IFC discrimination is slightly higher for vowels than for stops, although this is probably not significant.

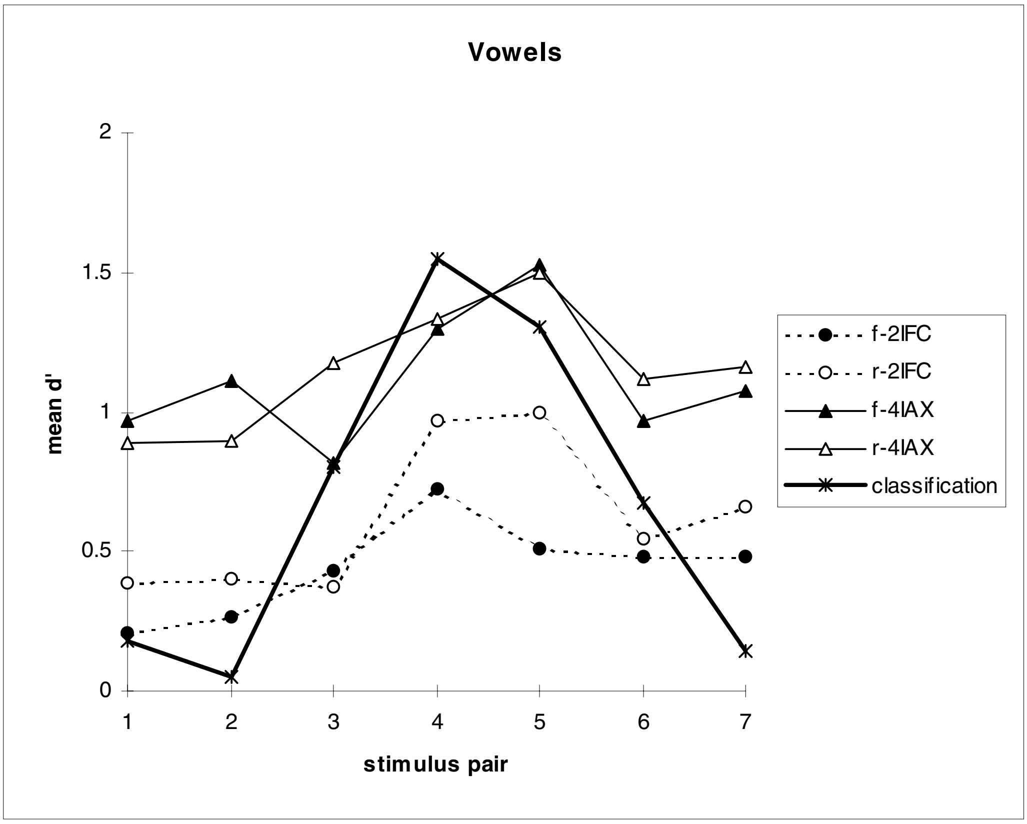


Figure 4.6. Classification and 2IFC and 4IAX discrimination results for the text vowels.

Statistical analysis

A two-way analysis of variance on the vowel data with factors Task (5 levels) and Stimulus (7 levels) demonstrated main effects of Task and Stimulus and a Task by Stimulus interaction effect (respectively F=24.34, p<.001; F=12.63, p<.001 and F=1.70, p=.023). Cell variance was over 11 subjects. A post hoc Tukey HSD test revealed that the main effect of Task implied that the mean d'-values for fixed and roving 4IAX were significantly higher than those for the other tasks. The post hoc analysis on the factor Stimulus demonstrated that the d'-scores at stimulus pairs 4 and 5 were significantly higher than the scores at the other stimulus pairs, which means that there was an overall discrimination peak at what appears to be the phoneme boundary location.

One-way analyses of variance per task were conducted, with Stimulus as independent variable, to reveal significant peaks in the discrimination results. Significant peaks were found for roving 2IFC and for classification, both at pairs 4 and 5 (roving 2IFC F=4.36, p=.001; Classification F=11.54, p<.001). There were no significant peaks in fixed 2IFC and fixed and roving 4IAX discrimination. Only in roving 2IFC was there a significant peak that coincided with the peak in the classification results. Thus roving 2IFC discrimination is more closely related to classification (which implies a higher degree of categorical perception) than are fixed 2IFC and fixed and roving 4IAX discrimination. However, even the roving 2IFC peak is lower than the classification peak. Similar results were found for the stop consonants. Apparently listeners use their phoneme representations in long term memory to discriminate stimuli, but they are less successful in using this strategy during 2IFC discrimination than they are in classification.

An analysis of the difference between the d'-scores for each stimulus pair revealed that fixed 2IFC was always significantly lower than the other tasks and that fixed and roving 4IAX were always higher except at pairs 4 and 5. At stimulus pair 4, Classification was highest and at pair 5 there was no significant difference between the d'-scores as a function of task. The analysis of the 4IAX data confirms what is shown in figure 4.6: within-category 4IAX discrimination was significantly better than predicted by classification. The within-category results in 2IFC discrimination are not significantly better than predicted by classification. These results support the previous conclusion: 2IFC discrimination is more closely related to classification than is 4IAX discrimination.

4.3.2.1 Differences among subjects

In the vowel experiment 71% of the total variance was explained by cell variance. Once again we divided the total subject group into four subgroups based on their 4IAX performance. Figures 4.7 to 4.10 show discrimination and classification performance in the highest (three subjects) and lowest quartiles (three subjects). It has to be noted that the subjects in these vowel quartiles are not identical to the subjects in the stop-consonant quartiles. Only in the lowest quartile was there an overlap of 2 of the 3 subjects. In the highest quartiles there was no overlap, which means that subjects differed in their performance as an effect of the stimuli or the phoneme contrast under investigation.

Separate analyses of variance on the vowel data per task with factor Quartiles (4 levels) as the independent variable revealed only a significant effect on roving 4IAX discrimination (F=4.46, p=.006). There were no significant differences in fixed 4IAX between quartiles, probably because of the large variation within the quartiles. Thus, like in the case of the stop-consonant results, there was no significant difference between the quartiles for the phoneme labelling tasks, classification and 2IFC.

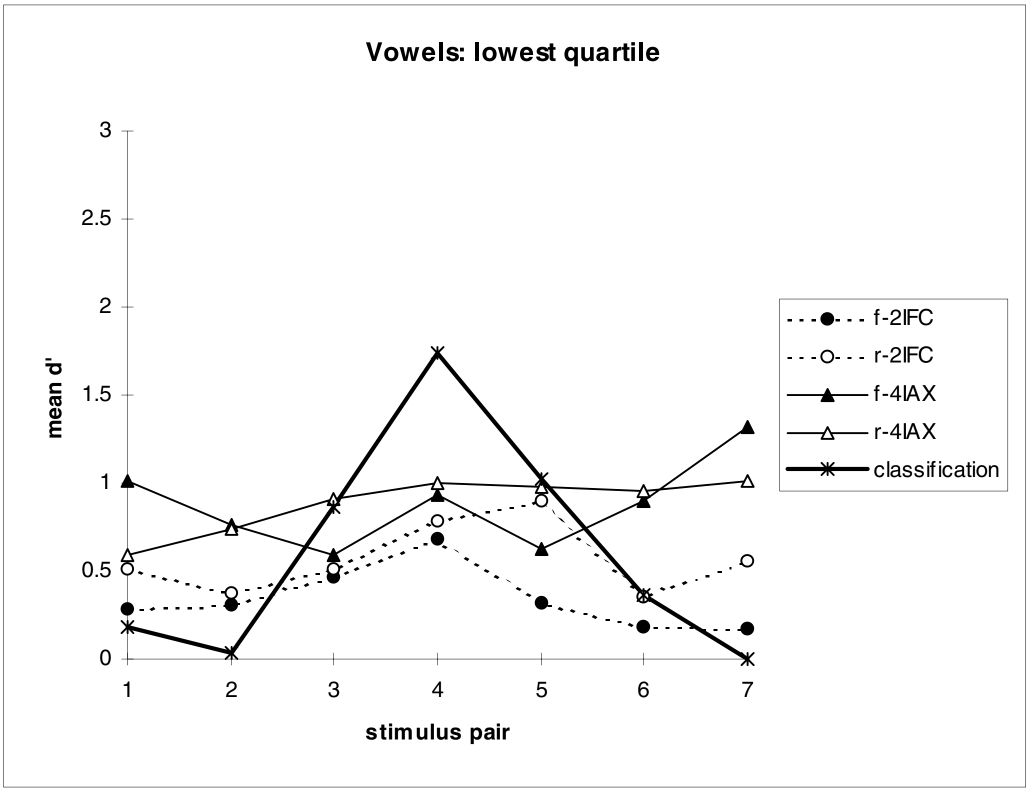


Figure 4.7 Vowel classification and 2IFC and 4IAX discrimination results of the lowest quartile (3 subjects).

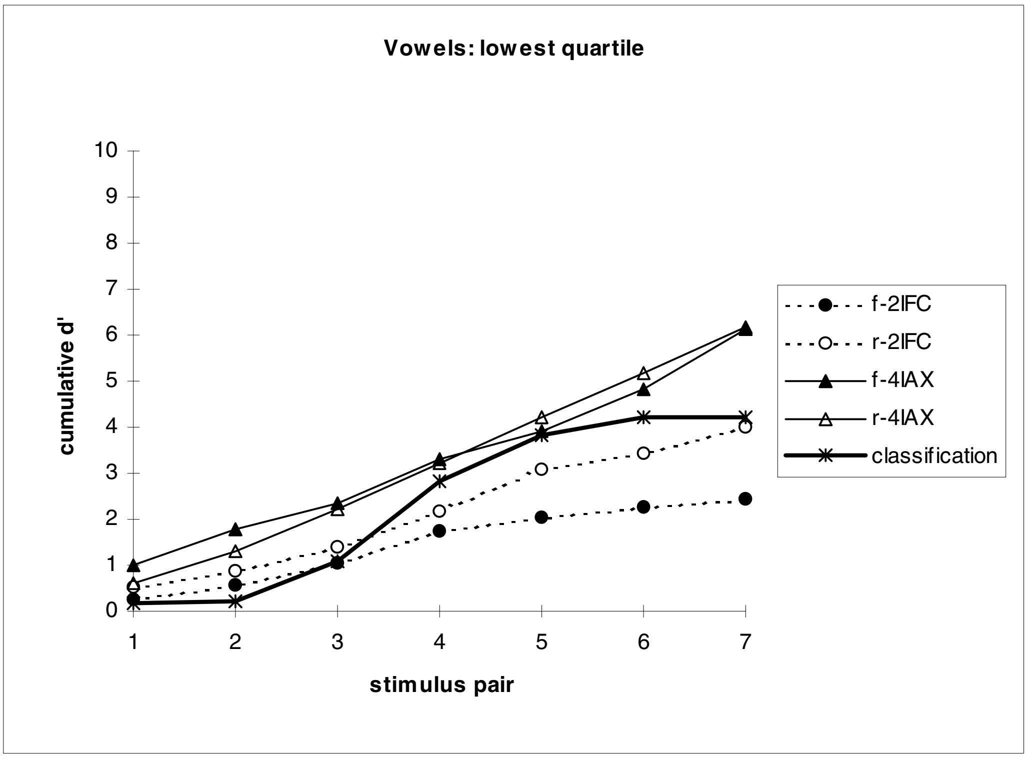


Figure 4.8 Cumulative vowel classification and discrimination results of the lowest quartile.

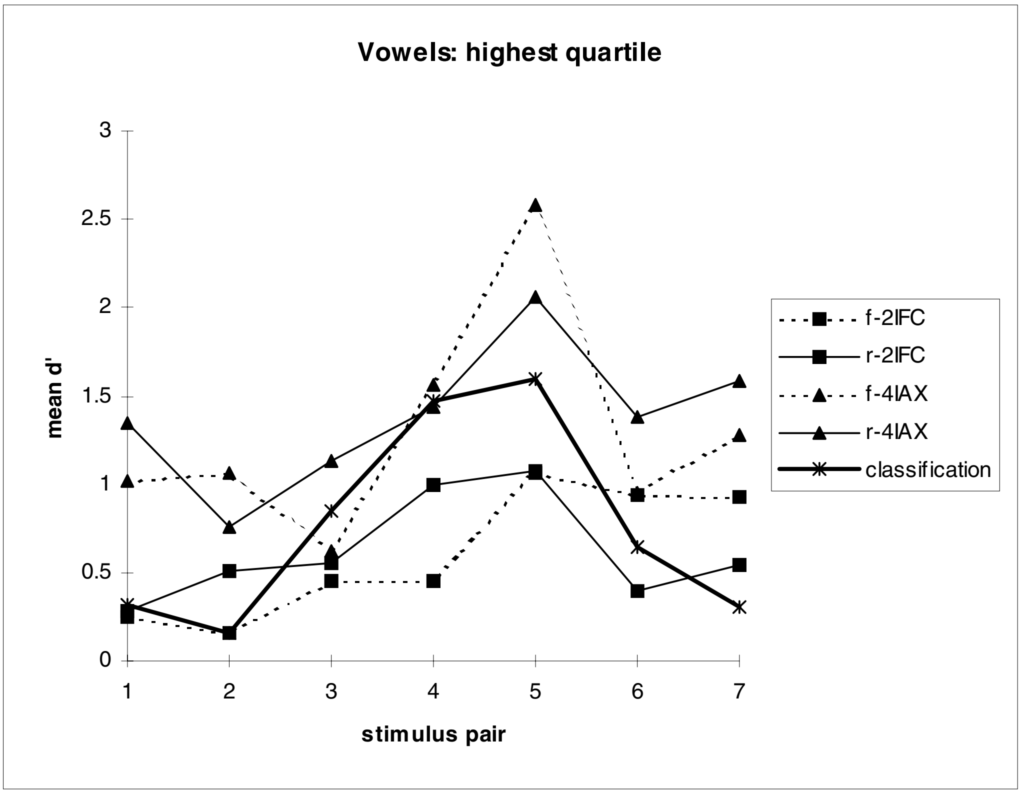


Figure 4.9 Vowel classification and 2IFC and 4IAX discrimination results of the highest quartile.

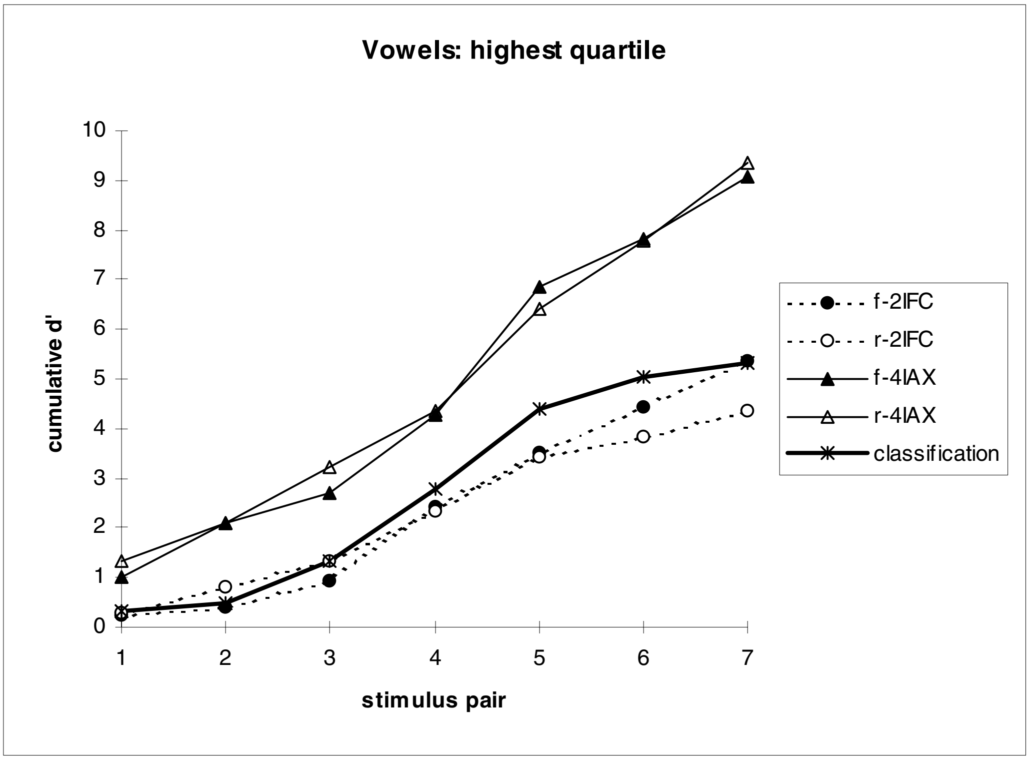


Figure 4.10 Cumulative vowel classification and discrimination results of the highest quartile (3 subjects).

Figure 4.7 displays the mean results of the lowest quartile. The individual data of these three subjects show that discrimination was always unrelated to classification, which is also the pattern in the average results. The individual results in the highest quartile (figure 4.9) were quite different from the average pattern. The most important individual effects are, for the first subject, a high peak in fixed 4IAX, which was at its maximum d' value (5.14), and for the second subject, a close relationship between roving 2IFC and classification. The results of the third subject showed a close relationship between fixed 2IFC and classification.

4.3.2.2 Discussion vowel results

The results clearly indicate that there is a difference in the degree of categorical perception for text vowels as a function of the discrimination task. 2IFC discrimination is more categorical (more related to classification) than 4IAX discrimination. The results are very similar to the stop-consonant results. Again we see that fixed 2IFC discrimination was lower overall than roving 2IFC discrimination. However, if we look at the individual vowel data we see that this was only true for 5 of the 11 subjects. Four subjects demonstrated similar fixed and roving 2IFC, and for two subjects fixed 2IFC was higher than roving 2IFC discrimination. It will be clear that there is a large variation within the subject group.

Another type of between-subject variation was found in the fixed 4IAX and roving 4IAX data. In some of the results fixed and roving 4IAX contained a marked discrimination peak, which coincided with the peak in classification. This peak was present in both the fixed and roving conditions for 4 of the 11 subjects. We found similar roving 4IAX results for some subjects with the word vowels in chapter 3, where, as an effect of a relatively long inter stimulus interval, some listeners tended to use more phoneme labelling, which caused scores to peak around the phoneme boundary.

As in the stop consonant findings, we see that roving 2IFC for vowels is worse than predicted by the labelling data. Again these results are very similar to those found by Schouten and Van Hessen (1992). Interestingly, the present vowel results imply that the text vowels used here are perceived differently from the word vowels. Even though we did not find any differences in 4I-oddity discrimination as an effect of vowel condition, we now see that 2IFC and 4IAX discrimination do reveal perceptual differences.

If we had used the model by Schouten and Van Hessen (1992), the roving 2IFC results for the text vowels would have shown a very close relationship between classification (similar to that found for the stop consonants). However, the recalculated 2IFC data for the word vowels would be better than predicted, which is exactly what Schouten and Van Hessen report for their ‘word’ vowel stimuli. This indicates that with the model by Schouten and van Hessen, the degree of categorical perception would be higher for the text vowels than for the word vowels.

In addition, 4IAX discrimination for the word vowels did not show any effect of phoneme labelling: there was no discrimination peak in the 4IAX word-vowel results. In the 4IAX results for text vowels we did find a discrimination peak for some subjects. We think this difference is caused by differences in complexity. Listeners will use their labelling strategy to discriminate stimuli especially when the spectral coding of the stimuli is more complex.

If we return to the question whether word vowels are perceived differently from text vowels, the answer is a qualified no. In the Schouten and Van Hessen model, text vowels are perceived more similarly to stop consonants than word vowels are. However, this interpretation is not really open to us since we chose to use the standard model by Macmillan et al. (1977). Therefore, the 'true' d' scores in the present study show that there is closer relationship between 2IFC and classification performance for word vowels than for text vowels (and also for stop consonants). For the 4IAX results the average functions are unrelated to classification in both vowel conditions, although there are some listeners that show a peak in 4IAX for the text vowels.

4.4 Discussion and conclusion

The two aims of this study were to see if the degree of categorical perception of stop consonants varies as a function of the discrimination task and whether the degree of categorical perception for stop consonants is higher than the degree of categorical perception for vowels.

Firstly, we found that the degree of categorical perception for stop consonants does vary as an effect of the discrimination task. If we follow the strict definition of Liberman et al. (1957), we have to conclude that the results show no categorical perception for stop consonants, because both 2IFC discrimination and 4IAX discrimination deviate from classification. Still, as we predicted, 2IFC performance is more closely related to classification performance than 4IAX discrimination. These results go against the general explanation of the high degree of categorical perception for stops proposed by Liberman et al. (1972). According to this explanation, the listener has no auditory image of the signal available, but only the output of a specialised processor that has stripped the signal of all sensory information and represents each phonetic segment categorically. The present findings show that listeners do have access to the auditory image of stop consonants. However, it is clear that the relative amount of auditory trace information that the listeners use during discrimination varies as a function of the task, the experimental context, and the listener.

Secondly, the present experiment has revealed that, as an effect of the use of two different discrimination tasks, the degree of categorical perception varies for vowels as well as for stop consonants. The effects of the different tasks are identical for both types of phoneme, thus stop consonant perception is not more categorical than vowel perception. The vowel results were not strictly categorical either, but again 2IFC performance was more closely related to classification performance than 4IAX discrimination. The main difference between the vowel and stop-consonant results was that within subject variance was 17 % higher in the vowel experiment. The resemblance between the stop and vowel results goes against the general view that these two types of phoneme are perceived in different perceptual modes, phonetic and nonphonetic (e.g. Ades, 1977; Fujisaki & Kawashima, 1970, 1971; Liberman, et al., 1972; Pisoni, 1973).

The results are not as clear as we might have hoped. From the earlier vowel experiments we draw the conclusion that, during discrimination of speech sounds, listeners use one of two possible strategies: they are either in a psychoacoustic mode and deal with the stimuli as essentially meaningless auditory events, or they are in a phonetic mode in which stimuli are interpreted as members of two possible phoneme classes (and auditory information has presumably been lost in the process and is unavailable to the discriminator).

The distinction between the two modes of perception was very clear in the previous experiments. All the evidence indicated that in 4I-oddity, listeners were not influenced by phoneme labels at all; they often failed to discriminate stimuli that they assigned to different phoneme categories during classification. There were good and poor discriminators, but neither group showed any relationship between discrimination and phoneme classification. In

2IFC, on the other hand, performance was largely, although not completely, determined by phoneme classification.

The present experiments with vowels and stop consonants reveal a different pattern. In the average data, 4IAX discrimination is quite similar to the earlier 4I-odity discrimination, and roving 2IFC is again determined largely by phoneme classification. Fixed 2IFC is different, however: performance is much lower than in classification and roving 2IFC (although, only for a subgroup of subjects), but there is a small peak at the phoneme boundary. In fixed 2IFC, where listeners know that they will be hearing the same two stimuli a large number of times, this knowledge does not seem to induce phoneme labelling at all: on the contrary, in roving 2IFC they are very much in the labelling mode, but in fixed 2IFC they are much more in the auditory mode. The labelling mode would lead to much more success, so there must be something about this particular series of experiments which causes listeners to abandon this strategy. However, we do not know what this might have been.

There is another difference with the earlier experiments: the highest quartile differs from the lowest quartile mainly around the phoneme boundary; this means that the difference between good and poor discriminators is not that the former are psychoacoustically better than the latter, but that they are better because they are capable of using labelling information when it is useful to them. Consequently, for some listeners, fixed and roving 4IAX discrimination were very similar to classification. This is true both of the vowel data and of the stop consonant data, which indicates that this result cannot be attributed to a difference in perception between these two types of phonemes.

Again, but this time with stop consonants, we see that the degree of categorical perception varies as a function of the discrimination task. However, the results of the experiments in the present chapter demonstrate that it is not just the task that determines the degree of categorical perception, but also the experimental context, the individual subjects, and perhaps the way subjects were motivated.

Chapter 5

The categorical perception hypothesis revised

5.1 Introduction

In this chapter a summary is given of the results described in the previous chapters. The summary is presented in separate sections that are related to the research questions of the first part of this study. These questions were:

- *Is there a difference in the degree of categorical perception of word vowels and text vowels?* (section 5.1.1)
- *Are there differences in the degree of categorical perception of vowels as a function of the discrimination task?* (section 5.1.2)
- *What is the effect of the duration of the inter-stimulus interval on discrimination and hence, on the degree of categorical perception of vowels?* (section 5.1.3)
- *Is there a difference in the degree of categorical perception of text vowels and stop consonants?* (section 5.1.4)

5.1.1 Vowel naturalness

In Chapter 2, we tested the hypothesis that text vowels would be perceived more categorically than word vowels. Two continua with eight vowel stimuli were presented to listeners in a 4I-oddity discrimination task and a classification task. The 4I-oddity task was chosen because it was expected that in this task listeners would have access to two listening strategies: phoneme labelling and a comparison of the auditory traces of the stimuli. The results showed that there were no differences in the degree of categorical perception between the two types of vowels; there was no indication of categorical perception at all, since observed discrimination was found to be completely unrelated to the classification data. A control experiment revealed that this was due to the nature of the 4I-oddity task, which took listeners out of a ‘phonetic mode’ and into a ‘psychoacoustic mode’.

5.1.2 Discrimination task

To investigate the influence of the type of discrimination paradigm on the degree of categorical perception, we compared performance on 5 different tasks within a single design. The tasks were 2IFC, AX, AXB, 4I-oddity and 4IAX discrimination. The results were as expected: there was wide variation in the relationship between discrimination and classification performance. In Chapter 2 it had already been demonstrated that discrimination with 4I-oddity was worse than predicted by the labelling data. For AX, AXB and 4IAX, observed discrimination was better than predicted. Furthermore, these tasks revealed within-subject variation in that some individual-subject results showed a discrimination peak that coincided with the phoneme boundary, whereas others did not. This indicates that the factor Listener also contributes to the variation in the degree of categorical perception. Finally, a 2IFC experiment involving the same vowel stimuli and the same listeners elicited the traditional categorical results.

5.1.3 Inter-stimulus interval

The duration of the temporal interval between stimuli in a discrimination task was varied in the experiments described in Chapter 3. The duration of the interval was 200 ms or 500 ms. It was found that, in general, discrimination improved as a consequence of the longer interval. However, it was not true that perception was always more categorical as an effect of the longer ISI. This turned out to be dependent on the discrimination task. Only in 4IAX was discrimination performance more categorical with the longer interval. This was not the case for 4I-oddity or 2IFC discrimination.

5.1.4 Stop consonants and vowels

In a final series of experiments, we compared categorical perception for stop consonants with vowels. Previous studies have demonstrated that the degree of categorical perception may vary as a function of the discrimination task for vowels, but that this effect is much smaller for stop consonants. However, with the text vowels used in our experiments, the effects were almost identical to those for the stops. It was found that 4IAX discrimination was less categorical than 2IFC discrimination for both phoneme classes. This is of particular importance, since for the first time it has been demonstrated, without giving listeners extensive training, that there is no overall phoneme labelling effect in stop-consonant discrimination.

5.2 A revised categorical perception hypothesis

In the previous chapters, references to Pisoni have occurred relatively frequently, mainly because several of his ideas have been tested in the present study. Pisoni applied the dual-process model to a variety of discrimination paradigms, and found that the degree to which perception is categorical depends, to only a small extent, on how much use can be made of auditory memory. He tested the effects of variation in stimulus duration, duration of inter-stimulus interval, and discrimination task (Pisoni, 1973, 1975; Pisoni & Glanzman, 1974; Pisoni & Lazarus, 1974). His findings never led him to conclude that categorical perception results were an artefact of the experimental design. Even Repp (1984), in his extensive review of the results obtained by Pisoni and colleagues, concludes that those results always remain fairly stable, despite all the attempts to influence the degree of categorical perception. The reason for this probably is that the results were never as clear-cut as ours are. For instance, stop-consonant perception always remained fairly categorical, despite the methodological manipulations. Even vowel perception did not vary a lot as an effect of task, discrimination always showing a peak that coincided with the phoneme boundary.

Since the dual-process model (Fujisaki & Kawashima, 1970, 1971) and its application to several experiments by Pisoni (1973, 1975), more attention has been given to an analysis of the experimental situation, including tasks factors, stimulus factors, and subject factors and their effects on categorical perception. We think that our findings raise further doubts on the laboratory phenomenon of categorical perception, as it turns out that almost all aspects of a particular experiment will influence the degree of categorical perception.

In conclusion, we will try to answer the main research question of the present study. This question was: *Is there an invariable relationship between discrimination and classification of a stimulus continuum between two phonemes?* We demonstrated that varying categorical perception results can be found for vowels. It has also become evident that the degree of categorical perception for vowels can be very high. In addition, it has been shown that the degree of categorical perception for stop consonants varies in a similar way. Various factors have turned out to influence the degree of categorical perception: discrimination task, stimulus naturalness, phoneme class, inter-stimulus interval, listener, and stimulus context. Therefore, we concur with Cutting (1982) who notes that categorical perception, defined as the relationship between experimental results within the confines of the standard classification-discrimination paradigm, is a “fickle phenomenon”.

The large degree of variation in the discrimination results, and therefore in the degree of categorical perception, indicates that the focus of categorical perception research should be mainly on classification and not on the relationship between discrimination and classification. Classification of speech stimuli requires a phonemic judgement, and decisions are based mainly on phonetic properties and features represented in long-term memory. Discrimination of speech stimuli, however, is not only based on phonetic representations; listeners can also use the differences between the auditory traces of the stimuli. We do not want to claim that listeners’ classification of isolated words on a continuum between two phonemes in a controlled laboratory experiment is analogous to the processes they use in everyday speech perception. However, classification is probably more closely related to these processes than is discrimination.

In chapter 1, it was noted that there has been a shift away from discrimination performance as a measure of speech processing and language development in recent infant studies (e.g. Jusczyk & Aslin 1995). It will be clear that our results support the notion that discrimination performance may not give a complete picture of infants' (or adults') language capacities. Results with older children are also in line with this idea. Sussman and Carney (1989) are among the very few who have tested discrimination *and* classification of a stimulus continuum between two phonemes with children. Their results demonstrate that children's labelling (/βA/ or /δA/) was adult-like at age 6, but there were significant age-related differences in discrimination performance, even for the oldest children in this study, who were 10 years old. This suggests that the children used different perceptual processes during classification and discrimination and that these processes develop independently.

Studies of children's classification of speech sounds have shown that, depending on the specific phonemes that were tested, children's performance reaches an adult level between 6 to 12 years of age (e.g. Kuijpers, 1996). In chapter 6 of this thesis we will describe the acquisition of one aspect of speech classification: the weighting of acoustic cues that signal phonemes. The experiment described in that chapter was motivated by the question of how children learn to classify speech sounds if there are no invariant cues that specify a certain phoneme category. It is assumed that adult listeners integrate various acoustic properties when they have to make a decision about a phoneme category, and that they assign different weights to these properties. How do children acquire these weighting schemes? We will study the weighting of certain cues for three different phoneme contrasts. The research on children's weighting of acoustic speech cues has mainly concentrated on one phoneme contrast, that of voiceless fricatives. Up to now, the results demonstrate that children are sensitive to detailed acoustic information in the speech signal. This supports the revised categorical perception hypothesis: the phonological system does not affect auditory sensitivity to relatively small acoustic changes in speech.

Chapter 6

The perceptual weighting of speech cues during childhood

6.1 Introduction

In the general introduction of this thesis it has already been said that the study of categorical perception is motivated by the fact that the acoustic signal lacks explicit linguistic structure. Nonetheless, under natural or experimental conditions, listeners manage to analyse the signal in such a way that discrete linguistic elements are recovered. Discovering how listeners analyse the speech signal to derive phonetic structure, despite the apparent lack of such structure in the physical signal, has been the most important goal in the field throughout its short history (e.g. Pisoni, 1985). It is assumed that adults' phonetic category judgements depend upon the perceptual integration of multiple acoustic properties spread across the spectrum and over time (e.g. Best et al., 1989; Kluender, Diehl & Wright, 1988). Several studies have demonstrated the perceptual integration of noise bursts and vocalic formant transitions for stop consonants (Dorman, Studdert-Kennedy & Raphael, 1977; Pols, 1979; Schouten and Pols, 1983) and of fricative noise and vocalic formant transitions for fricatives (Best, Morrongiello & Robson, 1981; Kunisaki and Fujisaki, 1977; Mann & Repp, 1980; Whalen, 1981). The perceptual weights adults assign to acoustic properties are the result of *learning* how phonetic information conveyed by each property varies across environments; this means that it is part of language development. For any given acoustic property, children have to learn to increase the weight assigned to that property in environments in which it is particularly informative, and to decrease its weight in environments in which it is less informative (Ohde & Haley, 1997; Nittrouer, 2000b; Nittrouer & Miller, 1997b).

6.2 Perceptual integration of speech cues by children

The developmental studies that have investigated perceptual integration have been concerned with two questions. Firstly, do children rely on the same acoustic cues as adult listeners when they have to make a decision about a phoneme category? And secondly, if so, do they assign the same weights to these cues during perceptual integration as adult listeners do? These two questions have been answered for stop consonant perception (see section 6.2.1), for fricative perception (section 6.2.2), and for vowel perception (section 6.2.3).

6.2.1 Cue integration for stop consonants

In the perception of initial stop-consonant place of articulation, the focus has been on age-related differences between the relative informativeness of the noise burst and the vocalic formant transitions. This relative informativeness was measured by deleting one cue from the onsets of words and determining to what extent children and adult listeners were still able to categorise the speech stimulus (Ohde et al., 1995; Ohde & Haley, 1997; Parnell & Amerman, 1978; Walley & Carrell, 1983). The results of Walley and Carrell (1983) show that the informativeness of the noise burst and the vocalic transitions is similar for adults and children of approximately 6 years and older: they can identify stops fairly well with only the burst or only transition information, but there is a tendency for all listeners to have higher identification scores with transition information than with burst information. In Ohde et al. (1995) classification performance was at a high level for all listeners (5- to 11-year-old children and adults) across most stimulus conditions: with or without the noise burst and moving or straight formant-transitions. The difference in the results between these two studies may be caused by differences between their stimulus materials. In Ohde et al. (1995), the noise-burst condition contained one glottal period without formant motion, whereas in Walley and Carrell (1983) it contained only the noise burst.

In Ohde and Haley (1997) and Parnell and Amerman (1978), a developmental effect of the vocalic transition property was obtained. The classification results of children (aged 3-5 years) improved relatively more in the formant-transition condition than did the adults' classification. Yet, in Ohde and Haley (1997) this age-related effect was limited to the perception of the [ɣ] stop consonant. The authors conclude that young children may perceptually weight formant transitions more heavily than adults, but only for a limited number of sound features.

An important drawback of some of these studies is that they may have used a method that is not very sensitive to age-related differences in the use of certain acoustic cues. For instance, in Ohde et al. (1995) there were marked ceiling effects for adults as well as for children, in that their classification scores reached 100% in all test conditions, even if one of the cues had been deleted. Ohde and Haley (1997), who were the only ones to test children as young as 3-4 years old, also reported ceiling effects in some stimulus conditions.

A more sensitive test was used by Hazan and Barrett (1999). They evaluated the role of the noise burst and the vocalic formant transitions for the stop-consonant contrast in "date-gate", by varying only one of the cues along a stimulus continuum and keeping the other cue at a neutral value (single cue condition). They also presented the stimuli in a 'combined cue' condition, in which the two cues were varied in harmony. In this study, listeners were 6- to 12-year-old children and adults. The stop consonant results demonstrated sharp labelling across the two single-cue conditions for all listener groups. Labelling functions were slightly steeper for the formant-transition cue than for the noise cue, but this difference was not significant.

The combination of the two stop consonant cues resulted in the steepest classification performance. However, comparison of the slopes of the functions across stimulus conditions is not really possible, because the axes are incommensurate. Since there were no age-related differences in the function gradients within the noise burst or the formant transition condition, the authors conclude that children's weighting of these stop consonant cues was adult-like at age six.

But although this design was more sensitive than the one in the previous studies, there still is a possibility that the single-cue conditions have not really been neutral with respect to the missing source of information. According to Mann and Repp (1980) and Smits, Ten Bosch and Collier (1996) the missing information may be encoded as supporting one alternative rather than another. It is unknown if this was the case in the study by Hazan and Barrett (1999). They only present the values of the function gradients. The location of the phoneme boundaries is not described, and no figures with classification functions are presented.

Children's and adults' perceptual weighting of two cues was addressed in a study by Morrongiello, Best, Clifton and Robson (1984), with experiments in which two acoustic cues were independently varied in a factorial design. In this study, the integration of speech cues was investigated in experiments that used the methods of information-integration theory (Anderson, 1982; Massaro, 1998). This approach is based on the independent manipulation of acoustic cues of a speech event in a factorial design (see also the adult studies by Kunisaki & Fujisaki, 1977; Mann & Repp, 1980; Whalen, 1981). Morrongiello et al. (1984) conducted a classification task in which 5-year-old children and adults had to report the presence or absence of the voiceless stop consonant in "say-stay". The stop was signalled by the silent gap duration between the offset of /s/ frication and the onset of voicing, and by the first formant transition onset. Two /σει-στει/-continua were presented, each containing the same systematic variation in the amount of silence (gap duration) following the /s/ noise.

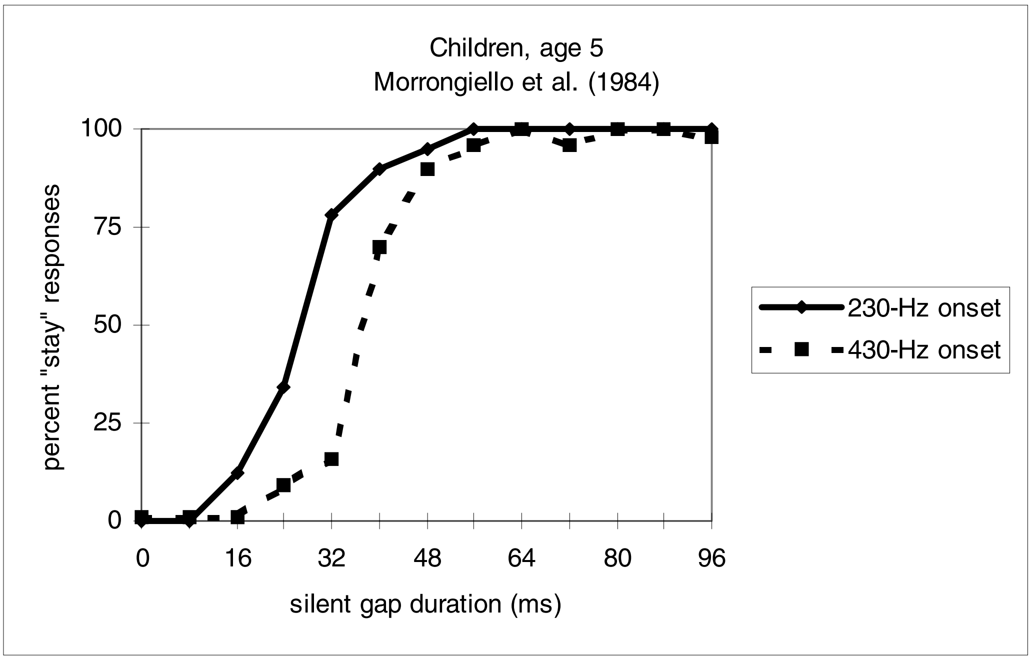


Figure 6.1. Classification results of 5-year-olds for /σει/-/στει/ syllables (figure reproduced from Morrongiello, Robson, Best & Clifton, 1984).

In one continuum, the vocalic portion had a 230-Hz F1 onset, which was expected to point towards the perception of /στεɩ/. In the other continuum a 430-Hz F1 onset was chosen that was expected to elicit /σεɩ/ as well as /στεɩ/ responses. In both continua the initial 40-ms transition was followed by a steady-state F1 at 611 Hz.

Figure 6.1 displays the classification results of the 5-year-old children. In an earlier study Best, Morrongiello and Robson (1981) had reported on the same experiment with adult listeners. The results are presented in figure 6.2. As can be seen by comparing figures 6.1 and 6.2, there is a shift of the phoneme boundary on the silent-gap duration continuum as an effect of the first formant onset. Furthermore, the /σεɩ-στεɩ/ classification functions are very steep in all age groups.

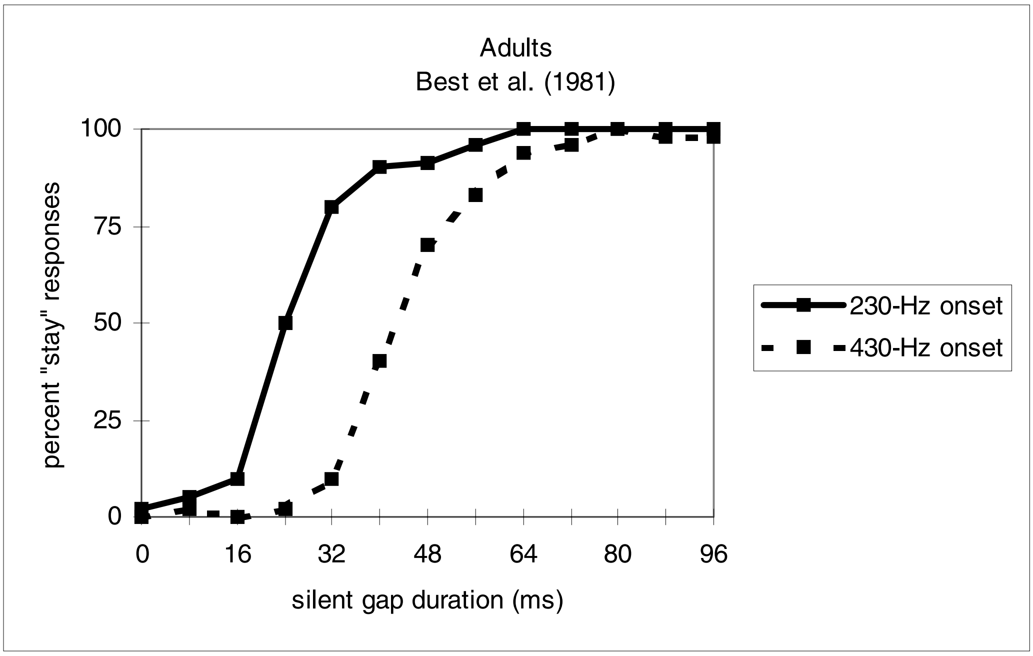


Figure 6.2. Classification results of adults for /σεɩ/-/στεɩ/ syllables (figure reproduced from Best, Morrongiello & Robson, 1981).

The figures also show that there is an age-related difference in the shift size of the functions: children needed less silence after the 430-Hz onset to hear /στεɩ/ than did adults. This difference seems to imply that children and adults differed in their relative weighting of the formant-transition cue. Since the children needed a shorter silent gap to disambiguate the 430-Hz onset, Morrongiello et al. (1984) conclude that the children weighted the formant transition cue relatively less heavily than the adult listeners did. This effect was not caused by absolute differences between children's and adults' perception of F1 onset information. In a control experiment it was shown that children and adults did not differ in their 'single-cue' classification of a 13-step stimulus continuum in which only the F1-onset was varied from 160 Hz to 611 Hz. Thus the 5-year-old children were equal to adults in their perceptual use of a 230- versus 430-Hz F1 onset as a single cue to a phonetic distinction. However, they deviated from adults in their weighting of the same acoustic information in a multiple-cued contrast.

These /σεɩ-στεɩ/ results were replicated with children of between 3 and 7 years old and adults by Nittrouer (1992) and Nittrouer, Crowther and Miller (1998). Their results show no age-related difference in steepness of functions, but a larger shift size between the adults' classification functions than between the children's functions. According to Nittrouer and colleagues these results indicate that children weighted the formant-transition cue more heavily than adult listeners, but their findings suggest that it is the other way around: the adult listeners weighted the vocalic formant transitions more heavily than the children (aged 4) did.

6.2.2 Cue integration for fricatives

The developmental trend found for perceptual integration of stop consonant cues was extended to fricatives by Nittrouer (1992, 1996a, 2000b). In Nittrouer and Studdert-Kennedy (1987), the weighting strategies for two cues of the "shoe-Sue" contrast were obtained from children (age 3;6 to 7 years) and from adults. The cues were the fricative-noise spectrum and the vocalic formant transitions. The stimuli consisted of two /Σu/ - /σu/ continua, each containing systematic variations in synthetic noise from /Σ/ to /σ/. The noise continuum consisted of nine aperiodic noise portions of 210 ms with centre frequencies ranging from 2.2 to 3.8 kHz in 200-Hz steps. In one continuum, the noise portions were concatenated with the vocalic portion from the /Συ/ vowel and in the other continuum with the vowel from the word /συ/. The results of this study, shown in figures 6.3 and 6.4, are similar to the findings presented in Nittrouer's more recent papers (e.g. Nittrouer, 1996a, 2000b; Nittrouer & Miller, 1997b).

From the results in these plots, Nittrouer and Studdert-Kennedy (1987) estimated to what extent labelling responses were based on the fricative-noise or the vocalic-transition cue. They took the slope of the classification function as an index of the weight assigned to the parameter represented on the abscissa, which in this case was the fricative-noise spectrum. The steeper the function, the more weight is assigned to the fricative-noise spectrum. The weight assigned to formant transitions is indexed by the separation between the functions. In all cue-weighting studies, the slope of the classification functions is interpreted as representing an index for the weight assigned to the cue that is represented by the stimulus continuum along the abscissa (e.g. Crowther & Mann, 1992; Escudero, 2000; Hazan & Barrett, 1999; Nittrouer & Studdert-Kennedy, 1987).

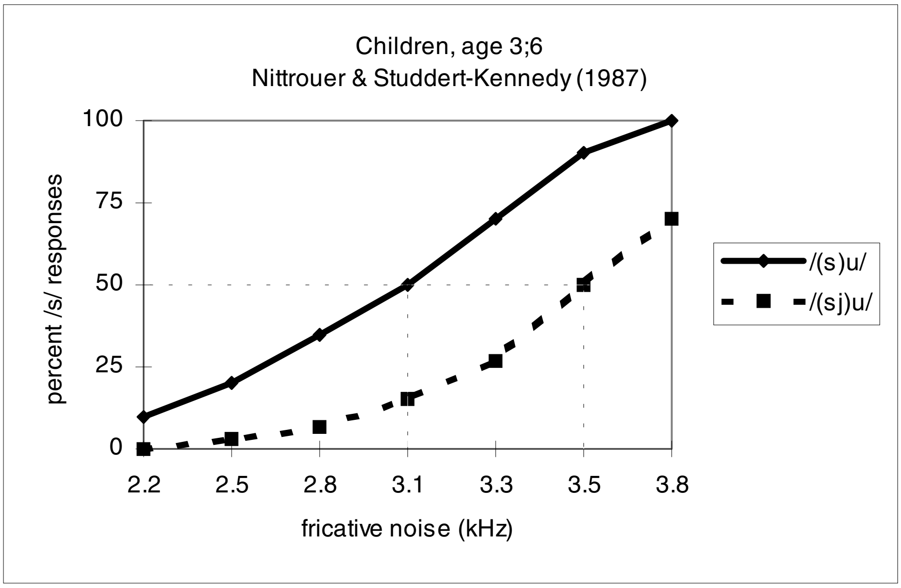


Figure 6.3. Classification results of three-year-olds for /Συ/-/συ/ syllables. Dotted lines indicate phoneme boundaries for each function and the separation between the two (figure reproduced from Nittrouer & Studdert-Kennedy, 1987).

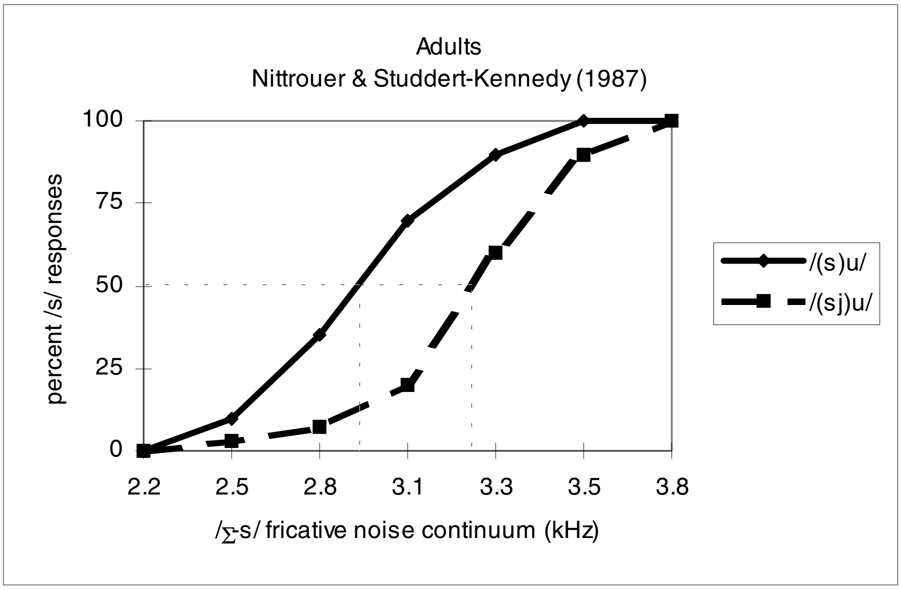


Figure 6.4. Classification results of adults for /Συ/-/συ/ syllables. Dotted lines indicate phoneme boundaries for each function and the separation between the two (figure reproduced from Nittrouer & Studdert-Kennedy, 1987).

In the more traditional classification studies, in which a single cue is varied, age-related differences are also measured by means of the slope of the classification functions. Usually, children's functions are shallower than those of adult listeners, which is said to indicate children's less consistent phoneme classification performance. Of course the weighting of an acoustic cue is correlated with classification performance: if children weight a cue less heavily than adult listeners do, it is expected that their classification function will be less steep than that of adults. In addition, it is expected that children's functions will be less steep because their responses will be less consistent than those of adult listeners due to more general cognitive factors, such as concentration and attention to the task.

The labelling functions in figures 6.3 and 6.4 are different in two aspects: 1) the children's functions are less steep than those of adults; 2) there is a larger separation between the two functions from the children. This indicates that formant-transition information was weighted more in children's than in adults' decisions on fricative identity and that the fricative-noise spectrum was weighted less (Nittrouer & Studdert-Kennedy, 1987). These weighting results have been replicated several times with the same /Σu/ - /σu/ combination (Nittrouer, 1992, 1996a, 2000b; Nittrouer & Miller, 1997b). The perceptual weighting of the noise cue and the vocalic transitions was also tested with fricatives in combination with other vowels: /Σι/ - /σι/, /ΣΑ/ - /σΑ/ (Nittrouer & Studdert-Kennedy, 1987; Nittrouer, 1992, 1996a, 2000b; Nittrouer & Miller, 1997a; and see Mayo (1999) for /Σo/ - /σo/). As a function of the vocalic context, the perceptual weight assigned to the vocalic transitions varied, but in all vowel contexts children (aged 3 to 7) demonstrated larger transition effects, and hence smaller noise effects, than adults.

Recently, Nittrouer found a similar pattern of results in a reversed manipulation of the fricative-noise and formant-transition cue (Nittrouer, 2000a). In this experiment, the stimuli again consisted of two /Σu/ - /σu/ continua, but now each continuum contained systematic variations in F2 formant transition from /Σu/ to /σu/. In one continuum, the noise portion was a slightly ambiguous /Σ/ noise and in the other continuum it was a slightly ambiguous /σ/

noise. The labelling results show that there is less separation between the children's (age 4) functions than between those of the adults. The children's classification functions look slightly steeper, but this is not significant. In general, these findings seem to confirm the previous age-related differences in weighting of these two cues: children's weighting of the noise cue is less than it is for adults, whereas the formant transition information is weighted relatively more by children than it is by adult listeners.

Nittrouer, Manning and Meyer (1993) termed the age-related differences in weighting of the /Σ-σ/ fricative cues the "Developmental Weighting Shift" or DWS. They propose that children's development of the appropriate weighting schemes is part of language learning. For fricatives this means that the vocalic transitions have more effect on young school-aged children's classification than on adults' classification performance, whereas the fricative noise has less effect on children's classification than on adults' classification. This effect changes as children gain experience with their native language.

In summary, it seems that with the labelling of /Σ/-/s/ compelling evidence exists that young children weight formant transitions more and noise spectra less than adults do. However, there is a problem concerning the characteristics of Nittrouer's fricative-vowel stimuli. In Nittrouer's initial experiments, the fricative-vowel stimuli always consisted of synthetic noise portions concatenated with natural vocalic portions (Nittrouer, 1992; Nittrouer & Studdert-Kennedy, 1987). As a consequence, it may be that the heavier weighting of formant transitions reflects the possibility that children find it more difficult to process information from synthetic than from natural speech components. This question was addressed in Nittrouer (1996a), Nittrouer and Miller (1997a) and in Nittrouer and Miller (1997b). In Nittrouer (1996a), the weighting of noise and formant transition cues was tested with two sets of fricative-vowel stimuli. One stimulus set consisted of synthetic noises with natural vocalic portions (hybrid stimuli) and the other stimulus set of synthetic noises with synthetic vocalic portions (synthetic stimuli). For the hybrid stimuli the same pattern of results was found as in previous studies: children weighted the formant transition cue more heavily than adults. However, this was not the case for the synthetic stimuli: there was no significant difference in the weighting of the formant transition cue between children (3-year-olds) and adults. According to Nittrouer (1996a) this absence of the transition effect is caused by the fact that only F2 was varied and not F3, as was the case with the natural stimuli. The idea that the F3 transition plays a substantial role is confirmed by Nittrouer & Miller (1997b). In experiment III of that study, again completely synthetic stimuli were used, but this time F2 *and* F3 were varied. The results show that now children did weight the transition information more heavily than adults did.

In conclusion, only one experiment so far has revealed DWS results with stimuli that did not contain partly synthetic and partly natural speech, but instead consisted of synthetic speech only. Therefore, it would be interesting to see if Nittrouer's results could be replicated, especially with completely natural stimuli.

In a recent study, Hazan & Barrett (1999) tried to replicate Nittrouer's /Σu/ - /σu/ results. Since the oldest children (age 7) in Nittrouer's studies still showed non-adult-like performance, Hazan and Barrett (1999) tested older children, aged 6 to 12, in order to see at what age their labelling results become adult-like. They used synthetic stimuli in which the fricative noise and formant transition were presented in single-cue conditions and in a combined-cue condition. However, the acoustic information of the vocalic-transition cue differed from Nittrouer's in that Hazan and Barrett only changed the onset of the F2 and not of the F3. In the single-cue conditions the other parameter was set to a neutral value.

It was found that the vocalic-transition functions of the youngest children were not steeper than those of the older listeners, and therefore that there was no evidence of greater use of vocalic transition information by the younger children. Hazan and Barrett (1999) explain the absence of an age-related difference in the perceptual use of the vocalic-transition cue by the difference between their and Nittrouer's formant transitions. Nittrouer (1996a) also failed to find any age-related transition effects with completely synthetic stimuli in which only the F2 transition varied. They further note that, according to Nittrouer, the shift away from formant transition cues is supposed to have arisen before the age of six. In their conclusion, Hazan and Barrett (1998) write that their results do not contradict the claims by Nittrouer and colleagues. But it has to be noted that the problem in their stop-consonant design (see section 6.2.1) may also apply to this experiment: it is not clear whether the neutral value of one of the cues in the 'single-cue condition' was truly a neutral one.

6.2.3 Cue integration for vowels

Concerning the effect of spectral acoustic cues on vowel perception, Ohde, Haley, and McMahon (1996) and Ohde and Haley (1997) tested if 3-, 4-, and 5- year-olds were more sensitive to formant transition cues than to steady-state formant cues. The stimuli were synthetic CV syllables composed of voiced stops followed by one of the vowels /ɩ-ʊ-A/ with either moving or straight formant transitions. In the moving-transition condition, the vowel portion started with formant transitions (F2 to F4) of 40 ms. The duration of the F1 transition varied across the different vowels. In the static formant condition, F2 to F4 remained constant at the onset frequency and only F1 movement was sustained. The authors found no evidence that formant transitions enhanced vowel identification in children. Children appeared to rely mainly on auditory cues provided by the steady states (Ohde et al., 1996; 1997). However, this vowel study produced even greater ceiling effects than their stop-consonant experiment (see section 6.2.1). In several stimulus conditions subjects responded with 90 to 100% correct identification. Young children understood the task and could identify the vowels at high levels, but the presentation of a CV syllable in which one of two cues is removed may not be the most sensitive instrument for measuring the relative importance of speech cues to phoneme perception.

6.2.4 Summary of children's perceptual integration of speech cues

The results described suggest that the relative importance of certain acoustic speech cues differs for young children, older children, and adults. In addition, there appear to be age-related differences between the relative weighting of certain acoustic cues. There is a trend in the data that implies that, in combination with spectral noise, formant transition information might be relatively more important, or weighted relatively more heavily, by children than by adults in making phonetic decisions. In the next section we will elaborate on some of the possible explanations for this finding.

6.3 Explaining children's perceptual integration of speech cues

The findings from the perceptual integration studies described suggest that adults and children divide their perceptual attention differently among the various components of the speech signal. The cause of these age-related differences has not been directly addressed in these studies. Only Nittrouer (1992) and Nittrouer and Crowther (1998) explicitly tested whether the development of the auditory system could explain their weighting results. In the next sections, we will elaborate on the auditory system explanation and on some other putative explanations of the differences between adults' and young children's perception, such as the development of the attention span, and of speech perception.

6.3.1 The development of speech perception

In the studies that have found age-related differences in the perceptual integration of speech cues it is suggested that this is dependent on linguistic development (Hazan & Barrett, 1999; Nittrouer, 1992, 1996a, 2000b; Ohde & Haley, 1997). The perceptual weights adults assign to acoustic properties are the result of learning how phonetic information conveyed by each property varies across environments, and children as old as seven years have not completely learned these informational-environmental relations. For any given acoustic property children have to learn to increase the weight assigned to that property in environments in which it is particularly informative (Nittrouer, 2000b). Several studies suggest that children's entry into language is mediated by meaning, and that the earliest unit of segmental contrast in child speech is not the phoneme, but the word or phrase (e.g. Bishop, 1997; Charles-Luce & Luce, 1990, 1995; Jusczyck, 1993; Liberman et al., 1974; Macken, 1992; Metsala, 1997; Stackhouse & Wells, 1997; Studdert-Kennedy, 1981; Vihman, 1996; Walley, 1993; Waterson, 1978). In these studies it is proposed that a 'global' representation is adequate during the early stages of acquisition because the number of words in an infant's lexicon is small. The representation of learned words in memory need not be based on details of the acoustic signal. As linguistic experience (and so, the lexicon) grows, representations that are more detailed are required in order to accommodate the increasing number of words to be stored. In Jusczyk's WRAPSA model (Word Recognition and Phonetic Structure Acquisition) it is assumed that only after the development of weighting schemes, language-specific phonetic decisions can be made (Jusczyk, 1993).

On the hypothesis that children have a more global listening strategy, Nittrouer has argued that the formant transition may be the acoustic correlate of a more 'global' cue. This correlation is motivated by the fact that a formant transition only emerges as a consequence of producing a stretch of speech larger than one phoneme, such as a CV syllable. According to her, this explains why children assign relatively more weight to the formant-transition cue than adult listeners do. Interestingly, one of Nittrouer's own series of experiments contradicts her 'global cue' explanation. Results by Morrongiello et al. (1984) and replications of these results by Nittrouer (1992) and Nittrouer, Crowther and Miller (1998) demonstrated that, if silent gap duration and F1 onset frequency are varied independently in a "say-stay" contrast, children (aged 4, 5 and 7) weight the transition cue less than adult listeners. The classification functions of all listeners were very steep, which means that children weighted the silent-gap duration in a way similar to adults.

Furthermore, findings in more recent studies seem to contradict the idea of 'global' lexical representations (e.g. Gerken, 1992; Jusczyk & Aslin, 1995; Swingley & Aslin, 2000). Swingley and Aslin (2000) demonstrated that infants of 18 and 23 months old recognised familiar words significantly more poorly when they were mispronounced than when they were pronounced correctly (e.g. vaby-baby, cur-car, tog-dog). These results suggest that the representations of words in infants' lexicons are well specified, even at a phonetic level.

In conclusion, we are inclined to agree with the hypothesis that age-related differences in weighting of acoustic speech cues are an effect of language development, but we do not believe that the weighting of formant-transition information is caused by a global strategy in speech perception.

6.3.2 The development of the auditory system

A factor that could play a role in the age-related differences in speech perception is the maturation of the auditory system. In general, it is said that children perform more poorly than adults on psycho-acoustic tasks (Allen & Wightman, 1992; Sussman, 1993; Sussman & Carney, 1989). Psycho-acoustic tasks contain both sensory and decision stages of processing; the ability of infants to discriminate speech sounds could be a reflection of the capacity of their sensory system.

Still, the greater importance assigned to formant-transition information as compared to noise cues may be due to difficulties in processing the latter rather than a global listening strategy. However, in Ohde and Haley it was found that even children as young as 3 years identified place of articulation for stops in the 'with burst / no formant transition' condition well above chance level. Nittrouer and colleagues explicitly tested the role of the auditory system on children's weighting schemes in two studies. Firstly, Nittrouer (1996a) demonstrated that children need larger acoustic differences in both fricative noise spectrum and F2 onset (formant transition) to discriminate /Συ-συ/ stimuli than adults do. Secondly, Nittrouer and Crowther (1998) studied the role of auditory sensitivity in a series of discrimination experiments with *nonspeech* stimuli. They obtained difference limens (DLs) from children (ages 5 and 7 years) and adults for three types for acoustic properties: F2 glide, F2 steady state, and silent gap. A pilot study showed that no subject (trained phoneticians as well as untrained listeners) perceived any of the nonspeech stimuli as speech. The results showed that children had larger DLs than adults for steady-state stimuli as well as for stimuli with transition-like properties. There were no age-related differences between the DLs for the temporal cue. The results of Nittrouer and Crowther (1998) indicate that differences in auditory sensitivity cannot explain the age-related differences in speech perception.

6.3.3 The development of the attention span

Other candidates for a list of factors affecting children's classification of speech signals include comprehension of instructions, motivation, attention to the task and task learning (Elfenbein, 1983; Sussman 1993). Elfenbein (1993) included these factors in her analysis and concluded that, although it is not possible to eliminate them as contributors to the developmental pattern observed, it is clear that they are not the sole contributors. In Nittrouer and Crowther (1998), results show that differences in attention to the task cannot explain age-related differences in perception. In this study there were no significant differences across the groups in perceiving the temporal cues (gap duration), but there were differences for other speech-like cues. According to Nittrouer and Crowther (1998), these results show that the measured differences cannot be attributed to the effects of procedural variables on performance. Firstly, children were able to perform the detection task itself equally well as adults, because in this one condition they showed similar results. Secondly, this finding shows that children's attention did not wane with testing because gap detection was the last task performed.

Furthermore, in Nittrouer's studies with fricative-vowel syllables, two test criteria were used (e.g. Nittrouer, 1992). Firstly, the best exemplars of /συ/ and /Συ/ were presented in a separate block between the test blocks and children had to identify these stimuli at least 90% correctly. In addition, children had to score 80% correct on these same stimuli during the test blocks, in which they were presented randomly with the other stimuli of the continuum. It seems reasonable to assume that children who meet these criteria have been concentrating during the task and also that they had no special difficulty performing it.

6.4 The present study

The perceptual integration studies tell us that children, like adults, do integrate various acoustic cues in making phonetic decisions. However, there appear to be age-related differences between the relative informativeness and weighting of certain cues. There is a general tendency for young children (3 to 7 years old) to rely relatively more on vocalic formant transitions than older children and adults do. From the preceding overview it will have become clear that research into the development of the perceptual integration of speech cues has been concentrated on stop consonant and fricative contrasts and that studies that report about children's perception of certain cues are not without methodological problems. One of the more robust findings seems to be the weighting shift of acoustic cues for fricatives as reported by Nittrouer and colleagues. However, in their weighting results there may be an effect of the quality of the stimulus material, which was partly synthetic and partly natural in all experiments except one.

The goal of the present study was to examine the perceptual weighting of various speech cues in a /Σ - s/, /π - κ/, and /α - A/ contrast. The aim was, firstly, to discover whether children of 4, 6, and 9 years old weight certain acoustic cues differently from adults and, secondly, whether it is possible to see if children give more weight to formant transitions than adults. The perceptual weighting was tested with a design similar to that of Nittrouer and colleagues, except that only completely natural stimuli were used. The cue-weighting design is an incomplete factorial design, since only one cue is varied along a stimulus continuum with the other cue remaining fixed at two prototype values. Ideally, a complete factorial design should be used with both cues varying along a continuum. However, this leads to a large number of stimuli, and hence to a much longer duration of the experiment. This is not a problem if only adult listeners are tested, but with children it increases the attrition rate as a consequence of boredom and lack of concentration. Therefore, we chose to keep the number of stimuli relatively low, as Nittrouer did, especially since we wanted to test cue weighting for three phoneme contrasts instead of one.

The research questions were as follows:

1) What are the effects of the fricative noise and the formant transition on fricative labelling by children and adults?
2) What are the effects of the noise burst and the formant transition on stop-consonant labelling by children and adults?
3) What are the effects of steady-state formant frequencies and vowel duration on vowel labelling by children and adults?

The /Σ/ - /σ/ contrast was chosen to see whether Nittrouer's results for American-English could be replicated (e.g. Nittrouer, 1992, 1996a, 2000b). The cues under investigation were fricative noise and formant transitions. It was expected that the /Σ/ - /σ/ classification results would show the same patterns as those of Nittrouer: the functions should be more widely separated and the slope should be less steep for young children than for adults.

We included the stop-consonant contrast in order to be able to compare fricative and stop-consonant labelling and to evaluate the relative use of noise and formant-transition cues across phonemic contrasts in which these cues may vary in duration and 'phonetic informativeness'. In fricatives the noise cues can be as long as 120 ms, but they may be as short as 10 ms in stop consonants. It is known from studies with adults that the formant transition is highly informative in making decisions about the place of articulation of stop consonants. However, adult listeners can also identify stops on the basis of the noise burst only. Dorman, Studdert-Kennedy and Raphael (1977) found an interaction between the relative importance of noise burst and transitions and the following vowel: in some vowel contexts the burst is more informative whereas in others the transitions have a more prominent role. However, Schouten and Pols (1983) showed that, in the case of initial voiceless plosives in Dutch, both burst and vocalic transition contain a great deal of information in any vocalic context. In addition, the findings of Van Wieringen (1995) demonstrate that the place of articulation of Dutch stops can be identified above change level from the transition cue only. On the basis of the results it is expected that adults will use both noise-burst and formant-transition cues to decide about stop consonant place of articulation. By contrast, the findings by Ohde and Haley (1997) and Parnell and Amerman (1978) lead us to expect that children will weight the formant transitions more heavily and the noise bursts less heavily for stop-consonant classification than adults.

By investigating vowel labelling we wanted to look into the relative weighting of a temporal and a spectral acoustic property. The temporal cue was vowel duration and the formant frequencies of the vowels represented the spectral cue. The vowels that were investigated were Dutch /α/ and /A/. The spectral difference between these vowels is relatively small: Koopmans-van Beinum (1973), Pols, Tromp and Plomp (1973), and Van Son and Pols (1990) report an F1 of about 640-690 Hz, and an F2 of about 1100-1300 Hz. The relatively small spectral contrast for /α/ and /A/ is supplemented by a difference in duration: the /α/ is a long vowel, its duration is about twice as long as that of the short vowel /A/ (Nooteboom, 1972). In Dutch both spectral and temporal cues seem to contribute equally to the discrimination of /α/ and /A/ (Nooteboom & Cohen, 1984; Van Heuven, Van Houten & De Vries, 1986). Clement and Wijnen (1994) and Kuijpers (1991) have found that 2- and 4-year-olds are able to realise the /α - A/ length contrast. Yet, the relative difference in length between the vowels was considerably smaller in the children's productions: the length ratio of /α/ to /A/ was 1.75 for adults, 1.54 for 4-year-olds, and 1.2 for 2-year-olds (Clement & Wijnen, 1994).

Although little has been reported with respect to development of vowel perception, it is generally agreed that the perceptual vowel system is fully acquired long before the consonant system (Fourcin, 1978). There are a few studies that have looked into children's perception of acoustic cues for vowels (Ohde & Haley, 1997; Ohde, Haley & McMahon, 1996). Concerning the effect of spectral acoustic cues on vowel perception, they demonstrated that 3-, 4-, and 5-year-olds and adult listeners relied equally on auditory cues provided by the static formant onset or the dynamic formant transition.

For the experimental manipulation of vowel duration or vowel formant frequencies, no specific predictions are offered in the developmental speech perception literature. On the basis of the results of Ohde et al. (1996, 1997) and the assumption that the perceptual vowel system is acquired at an early age, it is expected that the weighting schemes for vowels will be adult-like at an earlier age than the weighting schemes for consonants. According to Nittrouer, the weighting schemes for fricatives are adult-like when children are approximately 7 years old. If vowel perception develops earlier than fricative perception, it may be expected that the relative weighting of the vowel cues by children aged 7 or older will not differ from the weighting by adult listeners.

6.5 Method

6.5.1 Stimulus material

The stimuli used were word pairs in which two acoustic cues were manipulated that specified a phoneme contrast. In the consonant stimuli the noise and formant transitions were manipulated. In the vowel stimuli, formant frequencies and vowel duration were manipulated. The stimuli were generated by varying one cue in a way that resulted in a stimulus continuum between two unambiguous endpoints. For the other cue, binary values were used which were appropriate for one or the other target phoneme. The continua were formed via interpolation between the relative amplitudes of the spectral envelopes of the target phonemes. This resulted in very naturally sounding stimuli. The interpolation method was identical to the one used to generate the stimulus continua for the experiments in the previous chapters. In chapter 2, section 2.2.1, a more detailed description of the method can be found. All stimuli were digitised at a 44.1 kHz sampling rate.

In the next sections the procedure of interpolation and the stimulus characteristics are described in detail. The formant frequencies of the vowel portions were always analysed with linear predictive analysis (preemphasis, Hamming window, Levinson-Durbin algorithm). This procedure is described by Markel and Gray (1978).

6.5.1.1 Fricatives /ΣOκ - σOκ/ ‘plod’ and ‘sock’

The fricative stimuli consisted of a six-step noise continuum between /Σ/ and /σ/. Two settings were used for the /- Oκ/ portion that followed the fricative noise: in one setting the noise parts were concatenated with the /- Oκ/ portion from the word /ΣOκ/ and in the other setting the noise parts were concatenated with the /- Oκ/ portion from the word /σOκ/. Note that this means that the /- Oκ/ portions contained different formant transitions. The stimuli in the two settings will be referred to as /(Σ) Oκ/ and /(σ) Oκ/, respectively.

The original words were produced by a female native speaker of Dutch, as part of a list of words that contained six repetitions of /ΣOκ/ and /σOκ/ and filler words. The most natural and best matching /ΣOκ/ - /σOκ/ exemplars were chosen to serve as starting points for the stimuli.

Generation of the fricative-noise continuum

The continuum from /Σ/ to /σ/ was realised by interpolation between the relative amplitudes of the spectral envelopes of 190-ms noise parts from the original words. In the first step in generating the continuum, the signal was split into a source spectrum and a filter spectrum by means of cepstral deconvolution. The separation criterion was set at 2.63 ms in the quefrency domain. Next came an analysis of the spectral envelopes of the noise parts in terms of the phases and amplitudes of 100 spectral components between 80 and 8000 Hz, depending on spectral density. The spectral envelopes of the seven stimuli, obtained by means of six linear interpolation steps between each of the 100 pairs of spectral components, were then reconvolved with the original source spectrum of the /Σ/. The interpolation was always done in overlapping 19.18-ms time frames over the full length of the noise and the frame shift was 6.4 ms. In Nittrouer & Miller (1997a), the noise portions of a /Συ-συ/, and a /ΣΑ-σΑ/ stimulus continuum were also created by manipulation of the relative amplitudes of the spectral envelopes.

Description of the fricative stimuli

The fricative-noise portions were always 190 ms long. Figure 6.5 displays the spectral envelopes of the noise parts of stimulus 1 and stimulus 7. The major difference between the noise portions was a relatively high amplitude at roughly 2.5 kHz for stimulus 1 (the most /Σ/-like noise) and a relatively high amplitude at 4 kHz for stimulus 7, the most /s/-like noise. This is in line with the description of the spectral characteristics of these fricatives by Rietveld and Van Heuven (1997): more energy at relatively low frequencies for /Σ/ as compared to /s/. Figure 6.5 further displays the spectral envelope of the fricative noise in stimulus 4, which shows the relative amplitudes in the 2.5-kHz and 4-kHz regions, which are, due to interpolation, between those of stimuli 1 and 7.

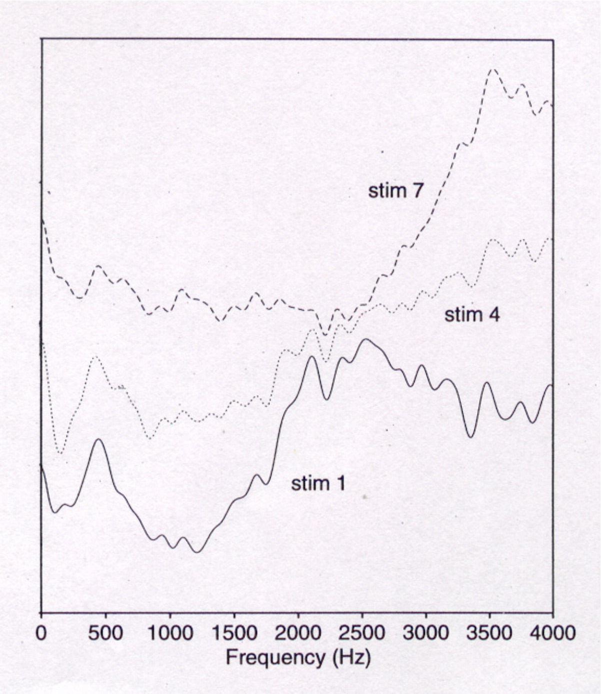


Figure 6.5. Spectral envelopes for the noise parts in stimuli 1, 4, 7 of the /ΣΟκ/ - /σΟκ/ continuum.

The duration of the **/(Σ) Oκ/**-stimuli was 480 ms, with a vowel duration of 130 ms. The formant transition onset frequencies were 220 Hz, 1787 Hz and 2566 Hz (first three formants). F2 fell through the entire vowel portion to a final frequency of 1046 Hz. The target positions of F1 and F3 were 460 Hz and 2340 Hz, respectively. The fundamental frequency fell from 228 Hz to 198 Hz.

The duration of the **/(σ) Oκ/** stimuli was 500 ms, with a vowel length of 120 ms. The onset frequencies of the first three formants were 276 Hz, 1328 Hz and 2778 Hz. F2 and F3 fell through the entire vowel portion to final frequencies of 1078 Hz and 2112 Hz. The final frequency of F1 was 424 Hz. The fundamental frequency fell from 229 Hz to 165 Hz.

6.5.1.2 Stop consonants /πOπ - κOπ/ 'doll' and 'cup'

The stop-consonant stimuli were composed of a six-step continuum synthesised by changing the noise burst from /π/ to /κ/. In one setting, these seven noise parts were concatenated with the /- Oπ/ portion from /πOπ/; this will be referred to as the /(π) Oπ/ setting. In the other setting, the seven noise segments were concatenated with the /- Oπ/ part from /κOπ/. This stimulus setting is called /(κ) Oπ/. The difference between the two settings is in the formant transitions. The original words were produced by a male native speaker of Dutch.

Generation of the stop-noise continuum

The continuum from /π/ to /κ/ was realised by interpolation between the relative amplitudes of the spectral envelopes of the original noise parts.

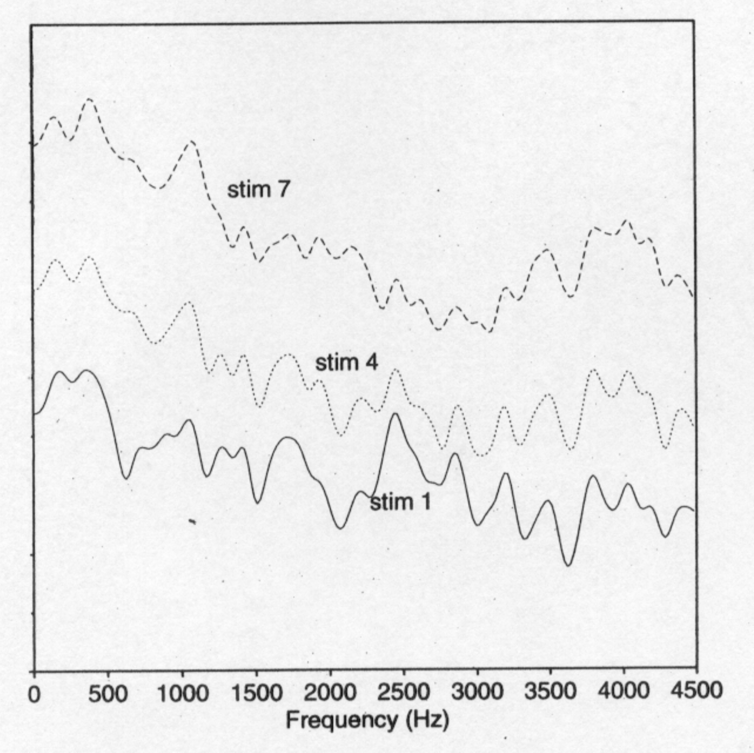


Figure 6.6. Spectral envelopes for the noise parts in stimuli 1, 4, 7 of the /πOπ/ - /κOπ/ continuum.

The spectral envelopes of the noise parts were analysed in terms of the phases and amplitudes of 80 spectral components. The time frame for the interpolation was 25.6 ms (frame shift 6.4 ms). The separation between the source and the filter was done with the criterion set at 0.50 ms in the quefrency domain. This procedure of interpolation between the amplitudes of the spectral envelopes was also used for similar stop-consonant stimuli in the developmental study by Walley & Carrell (1983).

Description of the stop-consonant stimuli

The noise bursts were always 30 ms long. The major differences between the bursts were the relative amplitudes of the frequency poles at approximately 2.7 kHz and 4.2 kHz, as shown in Figure 6.6. For the most /p/-like noise (stimulus 1), the amplitude was relatively high at 2.7 kHz, and for the most /k/-like noise (stimulus 7) it was relatively high at 4.2 kHz. Figure 6.6 also displays the spectral envelope of the noise burst in stimulus 4, which shows a relative amplitude in the 2.7-kHz region that is, due to interpolation, between that of stimuli 1 and 7.

The **/(π) Oπ/** -stimulus duration was 330 ms, with a vowel duration of 105 ms. The formant transition onset frequencies were 423 Hz, 805 Hz and 2518 Hz (first three formants). After 40 ms, the vowel target position was reached at 410 Hz, 899 Hz and 2553 Hz. The fundamental frequency was 133 Hz.

The total duration of **/(κ) Oπ/** was 300 ms, and the duration of the vowel was 100 ms. The onset frequencies of the first three formants were 391 Hz, 1032 Hz and 2070 Hz. After 45 ms the target position was reached at 428 Hz, 876 Hz and 2305 Hz. The fundamental frequency fell from 151 Hz to 142 Hz.

The formant frequency values reported here for Dutch /O/ are similar to those measured by Pols, Tromp, and Plomp (1973) and Van Son and Pols (1990).

6.5.1.3 Vowels /ζα κ - ζΑκ/ ‘shop’ and ‘bag’

The vowel stimuli consisted of a six-step continuum between the formant frequencies of /α / and /A/. In one stimulus setting, the vowels in the continuum had the appropriate duration for /α /. This stimulus setting will be referred to as /ζ (α) κ/. In the other stimulus setting the vowels in the continuum were shorter, because they had the appropriate duration for /A/. This setting is named /ζ (A) κ/. The original words were produced by a male native speaker of Dutch.

Generation of the vowel continuum

The continuum from /α / to /A/ was realised by interpolation between the relative amplitudes of the spectral envelopes of the vowels. The spectral envelopes of the original vowels were analysed in terms of the phases and amplitudes of up to 100 spectral components between 80 and 8000 Hz. The interpolation was always done in overlapping 25.6-ms time frames over the full length of the vowel (frame shift was 6.4 ms). For the vowels the separation criterion was set at 5.00 ms in the quefrency domain.

Description of the vowel stimuli

The duration of each **/ζα κ/** stimulus was 475 ms. Vowel duration was 200 ms. The values of the first three formants for /α / were 584 Hz, 1388 Hz, and 2392 Hz. The pitch was constant at 121 Hz. The duration of **/ζAκ/** was 435 ms and the duration of the vowel was 160 ms. The values of the first three formants for /A/ were 543 Hz, 1113 Hz, and 2305 Hz. The fundamental frequency rose from 122 to 139 Hz.

The formant frequencies are similar to those measured by Koopmans-van Beinum (1973), Pols, Tromp, and Plomp (1973), and Van Son and Pols (1990). The formants are marked by amplitude peaks in the spectral envelopes of the vowels in Figure 6.7. Stimulus 1 sounds like /α / and stimulus 7 is an /A/. It can be seen that the main difference between the vowels is in F2: for /A/ it has a lower frequency value than for /α /. The spectral envelope of stimulus 4 in the continuum is also given and shows relative amplitudes between those for stimuli 1 and 7.

The duration of the vowel in the stimuli was 200 ms in one continuum, and 160 ms in the other continuum. The difference in duration between the /α - A/ vowels was therefore 40 ms, which is less than the difference between the original vowels, which was 80 ms (220 ms versus 140 ms for /α / and /A/, respectively).

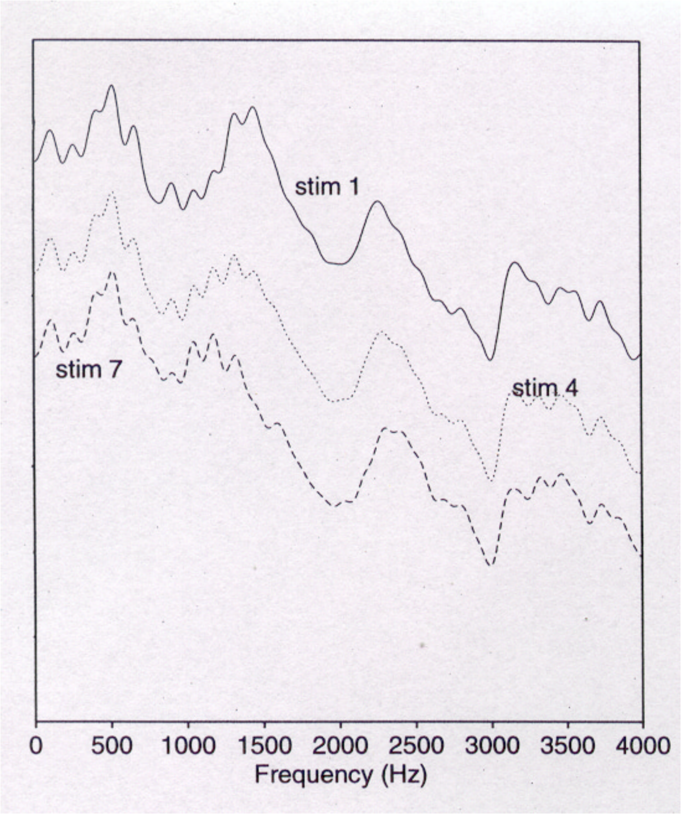


Figure 6.7. Spectral envelopes for the vocalic parts in stimuli 1, 4, 7 of the /ζα κ/ - /ζAκ/ continuum.

The values of 160 ms and 200 ms were chosen on the basis of phoneme judgements by four listeners (two phoneticians and two naive listeners). According to the pilot classification results, all stimuli in the /α - A/ continuum sounded like /ζAκ/, whenever vowel duration was less then 160 ms. On the other hand, stimuli with a vowel duration of > 200 ms were almost always classified as /ζα κ/. To avoid ceiling effects in the results as a consequence of the heavy weighting of the duration cue, i.e. only /zα k/ or /zAk/ responses, duration values were chosen that were more similar than the original duration values.

6.5.2 Subjects

The participants in this study were children of 4, 6, and 9 years old and adults. In each age group 11 listeners were tested, except for the age-4 group, in which 11 children completed the /ΣΟκ - σΟκ/ test, but only 10 children completed the /πΟπ - κΟπ/ and /ζα κ - ζΑκ/ tests. The children's age range was within 4 months below and 4 months above the designated age. They were recruited through a day care centre, a kindergarten, and an elementary school in Utrecht. The adults were students and employees at Utrecht University.

Selection and test criteria

To participate, children had to meet the following criteria:

- speech and language development at a normal pace,
- no siblings in speech therapy,
- no significant history of otitis media,
- no myringotomy tubes,
- no indices of otitis media within the 12 months prior to testing.

Once selected, each child participated in a Dutch speech discrimination pre-test (Auditieve Discriminatie Test, Crul & Peters, 1978). Only the first 15 of the 30 items of this standardised test were used. The children had to score at least 90% correct discrimination to be included in the classification experiment. The pre-test contained the stimulus words /σΟκ/ and /πΟπ/. The perception of /κ/, /α /, and /A/ was also tested, but with other words than the stimulus words. Unfortunately, the fricative /Σ/ did not occur in the test, because it appears in Dutch mainly in loan words, such as "chauffeur" (/ΣοэφΦρ/ driver), and "shampoo" (/эΣΑμπο/ shampoo). Apart from loan words, /Σ/ also occurs in diminutives, e.g. "jasje" (/эφΑΣ↔/ coat+dim.) and "huisje" (/эη↲ψΣ↔/ house+dim.). Although /Σ/ is a loan phoneme, it is regarded as part of the Dutch phonological system (Beers, 1997; Rietveld & Van Heuven, 1997).

Although /Σ/ was not tested in the discrimination pre-test, it was assumed that 4-year-olds have acquired this phoneme, which was confirmed by the fact that all children that participated in the classification test could produce the /Σ-s/ contrast in the test words.

Throughout the classification experiment, children and adults were expected to demonstrate at least 90% correct responses to the best exemplars of each response category. If post-test observations revealed that a subject had not done so, it was assumed that general attention had diminished, and the subject's data were discarded.

4-year-olds

A total of 34 children in the age-4 group were selected on the basis of the selection criteria. Five 4-year-olds were excluded, because they did not reach the criterion of the speech discrimination pre-test. This left 29 children who were available for the classification test.

For the /ΣΟκ - σΟκ/ contrast, the first 11 children tested (2 boys and 7 girls) completed the labelling test without any problem. Their mean age was 4 years.

In the /πOπ - κOπ/ contrast, 29 children were tested, but 4 children could not be motivated to participate in the last test session. Of the 25 children that completed the test only 10 children met the test criterion. Thus 10 children (5 boys and 5 girls) provided the data for this phoneme contrast. Their mean age was 3 years and 10 months.

In the /ζα κ - ζΑκ/ test, 10 children were successfully tested. All 29 children began testing, but 4 did not finish the last test blocks. Only 10 of the 25 children that completed the test met the test criterion. The boy/girl ratio was 2/8 and their mean age was 4 years and 2 months.

Note that only two 4-year-olds (2 girls) completed the tests of all three phoneme contrasts. Thus, the results for the three contrasts consist largely of judgements from different 4-year-old children, whereas in the other age groups the same children completed all tests.

6-year-olds

On the basis of the selection criteria 12 children of age 6 were selected. They all passed the speech discrimination pre-test and met the test criterion. However, one 6-year-old child did not finish all the tests, due to lasting illness. Since there were no other children in this age group that met the selection criteria, in total 11 children of 6 years old completed the tests. The group consisted of 8 boys and 3 girls, their mean age was 5 years and 11 months.

9-year-olds

Twelve children of age 9 were selected on the basis of the selection criteria. They all completed the speech discrimination pre-test and the labelling test without any problem, except one child who got sick for an undefined time. This left 11 successfully tested children in this age group. The mean age of 9-year-olds was exactly 9;0 and the group was composed of 6 boys and 5 girls.

Adults

As might be expected, the 11 adult participants were not faced with any problem. The adults were between 22 and 36 years of age. The male/female ratio was 5/6.

6.5.3 Design

The stimuli were presented in three sessions: one session for each phoneme contrast. The order of the sessions was alternated across subjects. There were three test blocks within each session. A test block contained 28 trials: 2 repetitions x 7 stimuli x 2 settings). In total each stimulus was presented 6 times. The total number of test trials in one session was 84. The order of the stimulus settings (the two different formant-transition cues) was alternated across subjects.

Figure 6.8 gives an overview of the training and test blocks within one session. The order of stimulus presentation within blocks 1 and 2 differed, but the order in blocks 1 and 3 was identical. The first test block was repeated to evaluate test-retest reliability. There were two versions of the order of stimulus presentation. In between test blocks, a block of ten best exemplars was presented (5 repetitions of each response category). There were two versions of the order of best-exemplars presentation.

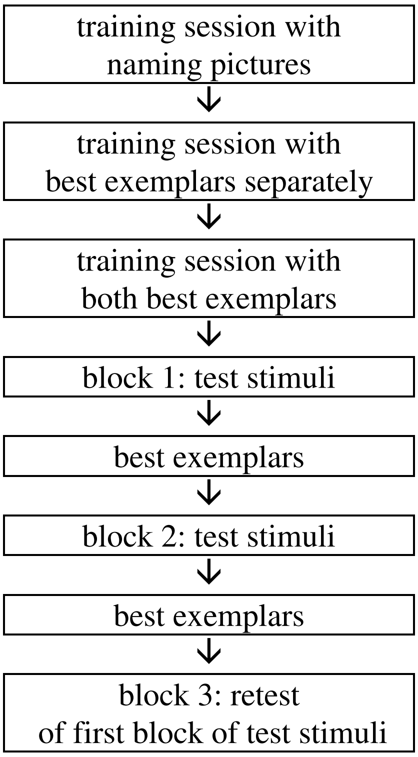


Figure 6.8. An overview of the experimental procedure that was used with the children and the adults. Adult testing started with training of both best exemplars.

6.5.4 Procedure

Testing with children

The children were tested in a quiet room within the school building. Testing was accomplished in 3 sessions. The first session contained the pre-test and the classification of one phoneme contrast, and this session lasted approximately 25 minutes. The other two sessions only contained the testing of one phoneme contrast and took approximately 15 minutes each.

The classification test was presented via a personal computer and the child heard the stimuli via headphones. The personal computer was a Pentium with Sound Blaster 16 sound card from Creative Labs. The headset was a Beyerdynamic TD770. Stimulus presentation was controlled by a Java program, which also generated a data file. Each stimulus was preceded by a 200 Hz pure tone that functioned as an attention signal. Children responded by pointing at the picture of the stimulus and saying what was heard. There was no response time limit. If a discrepancy was detected between what was said and what was indicated, the experimenter asked for clarification. The experimenter monitored stimuli over headphones to ensure that there were no equipment failures and to check the subjects' responses during testing.

After a short introduction of the test equipment, training started with the child learning to associate the pictures and the test words (see Appendix A). The experimenter asked the child to repeat the word she said and to point to the picture that represented the word. After five correct trials the game was reversed: the child had to name one of the pictures and the experimenter repeated the word and pointed. Sometimes the experimenter deliberately gave the wrong answer to see if the child would correct her. After a few trials the experimenter took over the game and this phase ended after the child again had given five correct answers.

In the second training phase, one picture was placed in front of the child and the best exemplar of that category was presented over the headphones. After this had been completed for one response category, the picture was removed and the procedure was repeated with the other response category. In the third training phase, both pictures were positioned in front of the child at the same time, and the child was trained to point to the appropriate picture and to repeat the stimulus (the experimenter entered the responses into the computer). The best exemplars of each response category were presented 10 times each, in two blocks of five each. The experimenter provided help during the first block, but not during the second block. This second block of best exemplars served to evaluate whether each child could respond correctly to 9 of 10 stimuli, a criterion for participating in the test.

Finally, testing with all stimuli in the set was conducted. After each block there was a short pause in which the child received a piece of a jigsaw puzzle or was allowed to turn over some memory cards. When the session was completed, children were given a sticker and after the three sessions they received a little toy.

Testing with adults

Classification testing with adults also involved a forced choice between two alternatives. Adults were tested in the experimenter's office in order to create a test environment comparable to the one at the children's school.

Like the children, adults responded by pointing to the picture that represented the stimulus and by repeating the word. There was no response time limit. The test procedure differed from that of the children in that pre-test training consisted of just one block with the best exemplars of the two categories and in that no games were played between the test blocks. Responses to the best exemplars of each phoneme category had to be 90% accurate or better.

6.6 Results

6.6.1 Statistical analysis

The results for the phoneme contrasts are presented in two classification functions for each age group. The classification functions represent the participants' mean proportions for one of the response labels for each stimulus. The slopes of the classification functions were analysed to see if there were any age-related differences in the effect of the cue represented on the abscissa. To determine age-related differences of the effect of the binary cue, the separation between the functions was measured.

Comparison of slopes

Both the slope (or function gradient), and the function separation serve as indices of the extent to which response decisions are based on each of the acoustic cues. The slope of the function is taken as an index of the weight assigned to the cue that is represented on the abscissa (notably the fricative-noise, noise-burst, and vocalic formant-frequency continuum).

To calculate the slope, Linear Regression analysis was used. The proportions were transformed into z-scores using the cumulative normal distribution. If the proportions for successive stimuli did not differ by more than .05, they were interpreted as showing asymptotic performance and hence, were left out of the analysis. Next, for each participant a Linear Regression analysis was carried out, with Stimulus as the independent variable and the z-score as the dependent variable. The slope coefficients (the b values) were analysed with a two-way analysis of variance with factors Age (4) and Transition/Duration (2).

Comparison of shift size

The effect of the binary cue was determined by measuring age-related differences in the separation between the two classification functions. In Nittrouer's studies (e.g. Nittrouer, 1992, 1996a, 2000b), the size of the weighting shift is determined by measuring the distance between the binary transition functions at the phoneme boundary. The phoneme boundary is defined as the place along each function at which categorisation performance is at chance level (50%). This procedure was not adopted here, since a weighting shift typically shows up over a range of boundary steps, not just at one or two. A more complete measure of the shift can be obtained by measuring the shift across a range of stimulus steps. Consequently, the method of Pitt & Samuel (1993) was used, in which an ambiguous region is defined and the mean classification proportions of each stimulus in this range are subtracted. In the current study the whole stimulus range was chosen as the ambiguous region. This choice was motivated by the way the stimuli were generated: due to the independent manipulation of the two cues, both stimulus continua consist partly of stimuli with conflicting acoustic information.

To measure the separation between the curves, the classification proportions were analysed in two stages. First, each proportion was transformed into an arcsine[2] score (Studebaker, 1985). Second, the area between the curves was measured by subtracting the arcsine scores at stimuli 1 to 7 of one continuum from their counterparts in the other continuum. A two-way analysis of variance was performed on the arcsine-difference scores for each phoneme contrast to test the effect of Age (4) and Stimulus (7).

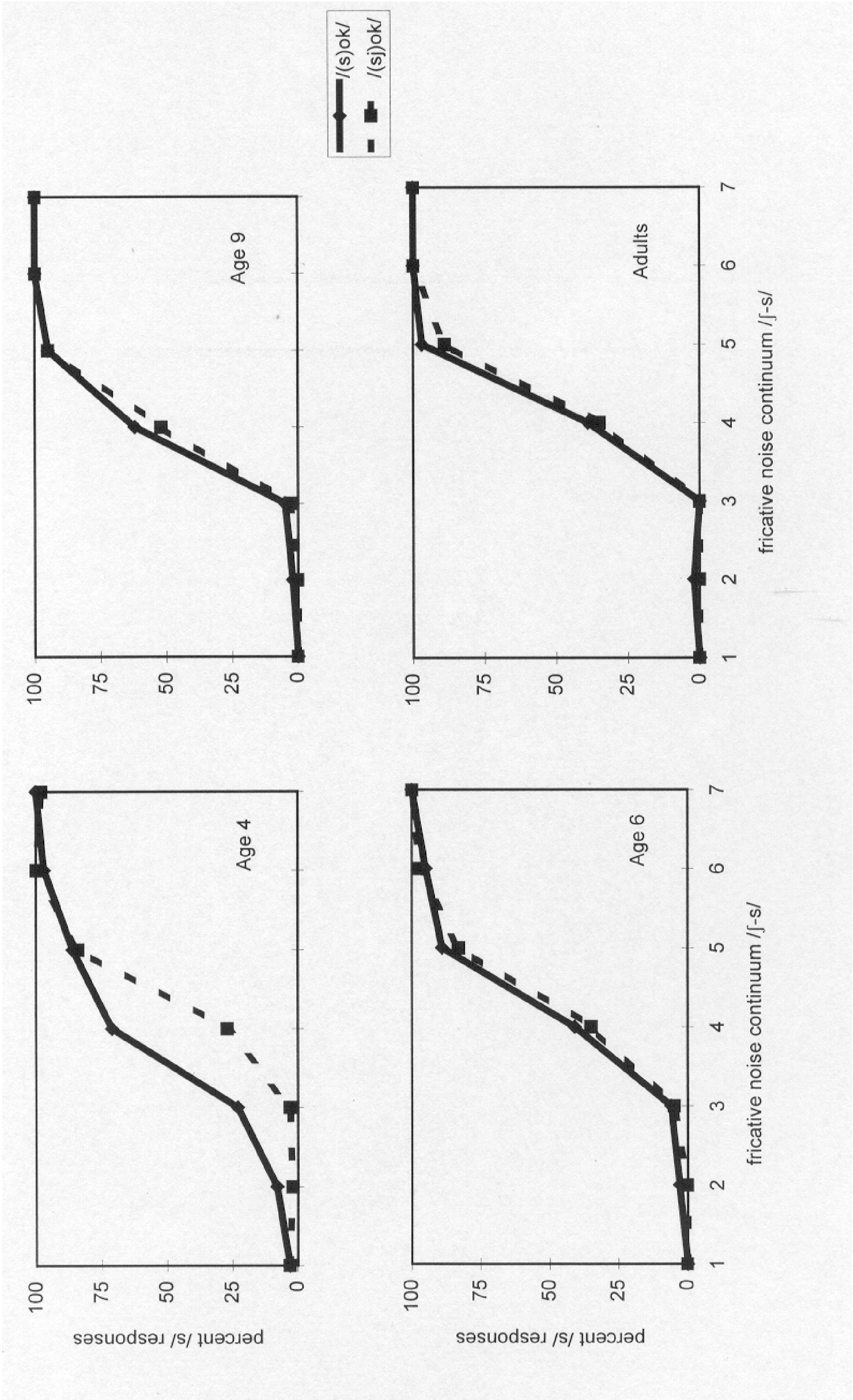
6.6.2 Fricative results

In figure 6.9 the results for the /ΣOκ/ - /σOκ/ contrast are presented. The participants' /s/ response proportions were calculated for each stimulus along the continuum. Within each age group 66 judgements were available per stimulus (6 presentations x 11 participants). In tables 1 and 2 of Appendix B, the mean proportions of the two functions are listed, along with the standard deviations.

In the analysis of variance on the /ΣOκ/ - /σOκ/ responses, the main effects of Age (4), Transition (2), Stimulus within Transition (7), and Test (3) were computed. There was no significant effect of Test, which indicates that listeners performed similarly in the three successive test blocks within the test session.

---

[2] To allow the arcsine transformation to be carried out, p-values of zero and unity were replaced by, respectively, P=.0009 and P=.9991 (see Kirk, 1982).


Age 4
Age 9
Age 6
Adults
percent /s/ responses
0
25
50
75
100
1
2
3
4
5
6
7
fricative noise continuum /ʃ-s/
/(s)ok/
/(sj)ok/

All the other factors had significant effects on the labelling responses: Age (F=7.28, p<0.001), Transition (F=25.33, p<0.001) and Stimulus within Transition (F=996.57, p<0.001). A post-hoc Tukey HSD test on Age revealed no pairwise differences between the age groups.

A comparison of the mean scores for the two series of stimuli (the effect of Transition) indicates that listeners, regardless of age, gave more /s/ responses when the transitions were appropriate to a preceding /σ/. There were significant Age x Transition and Age x Stimulus within Transition interactions (F=6.22, p<.001 and F=3.88, p<.001). The age-related differences were examined more closely by means of comparisons across age groups for the functions' slopes and shift sizes, which are listed in table 6.1.

The relative weighting of the fricative noise cue was determined by analysing the slope values for each age group. A two-way analysis of variance (Age x Transition) performed on individual slope values exhibited a significant main effect of Age (F=10.90, p<0.001). A post-hoc Tukey HSD test revealed that slope values were significantly lower for the 4-year-olds and the 6-year-olds.

Table 6.1. Mean slope and shift size of the /Σ Oκ/ - /σ Oκ/ classification functions for each age group.

	Slope	Shift size
Age 4	1.10	.23
Age 6	1.26	.05
Age 9	1.84	.05
Adults	2.09	.05

The separation between the classification functions, presented as 'Shift size' in Table 6.1 indicates the effect of the formant transition cue. A two-way analysis of variance with factor Age and factor Stimulus was performed on the arcsine-difference scores and revealed significant effects for both factors (F=4.43, p=.005 and F=7.43, p<.001). There also was a significant Age by Stimulus interaction, F=3.25, p<.001. A post-hoc Tukey HSD test for the Age effect revealed that the separation between the curves of the youngest children was significantly larger than that of the other age groups. The greater separation between the functions indicates that the 4-year-old children weighted the formant transition cue more heavily than the 6- and 9-year-old children and the adults. A post hoc Tukey test with Stimulus showed that the separation between the functions was largest at stimuli 3 and 4.

The results demonstrate that the 4-year-olds showed less weighting of the fricative noise cue and heavier weighting of the formant-transition cue than the older listeners, but only for the ambiguous stimuli in the continua. The 6- and 9-year-old children and adults did not use the information in the formant-transition cue for the ambiguous stimuli at all, but based their labelling decisions entirely on the information in the noise spectra.

The function gradients of the 4- and 6-year-olds were shallower than those of the 9-year-olds and the adults. This means that there is no perfect reciprocal relationship between the weighting of the two cues: the 6-year-olds show no use of the formant transition cue, but their function gradient is as shallow as the one of the 4-year-olds. Apparently, the 6-year-olds have already learned to base their fricative decisions mainly on the fricative noise, but their use of this cue is not adult-like yet.

6.6.3 Stop-consonant results

Figure 6.10 shows the mean percent /πOπ/ - /κOπ/ responses per age group. Only 10 four-year-olds completed the whole test. Thus, there were 60 judgements per stimulus for the age-4 group and 66 judgements per stimulus in each older age group. In tables 1 and 2 of Appendix C, the mean proportions of the two functions are listed, along with the standard deviations. In figure 6.10, it is illustrated that for the adult listeners the four endpoints of the /p-k/ contrast were clear, unambiguous tokens. For the other age groups, the children, only the best exemplars were unambiguous. The labelling results vary across the age groups, with marked differences in the slopes and shifts of the classification functions. In the analysis of variance on the /πOπ/ - /κOπ/ data, the main effects of Age (4), Transition (2), Stimulus within Transition (7), and Test (3) were computed.

The analysis revealed main effects of Age (F=26.69, p<.001), Transition (F=734.22, p<.001) and Stimulus within Transition (F=267.87, p< 0.001). There was also a significant Age x Transition and Age x Stimulus within Transition interaction (F=17.81, p<.001 and F=5.65, p<.001).

A post-hoc Tukey HSD test with factor Age showed that, across both stop-consonant transition conditions, the 4-year-olds gave more /k/ responses. This can be seen in figure 6.10: the 4-year-olds give 50% /k/ responses to the first stimulus of the /(κ) Oπ/ continuum, whereas this proportion is much lower for the other age groups. A comparison of the mean scores for Transition revealed that, when the transitions were appropriate for a preceding /k/ rather than a /p/, listeners gave more /k/ responses. The age-related differences were examined more closely by means of comparisons across age groups for the functions' slopes and shift sizes.

In table 6.2 the mean slope values and shift size values per age group are displayed. Since the effect of Transition is significant, the slope values for the two functions are presented separately. The two-way analysis of variance (Age x Transition) on the individual slope values revealed a significant effect of Age (F=17.85, p<.001) and Transition (F=5.84, p=.018), but there was no significant interaction between these two factors.

Table 6.2. Mean slope and shift size of the /πOπ/ - /κOπ/ classification functions for each age group.

	Slope		Shift size
	/πOπ/	/κOπ/	
Age 4	.79	.59	1.01
Age 6	.88	.87	.61
Age 9	1.19	1.13	.50
Adults	1.93	1.29	.66

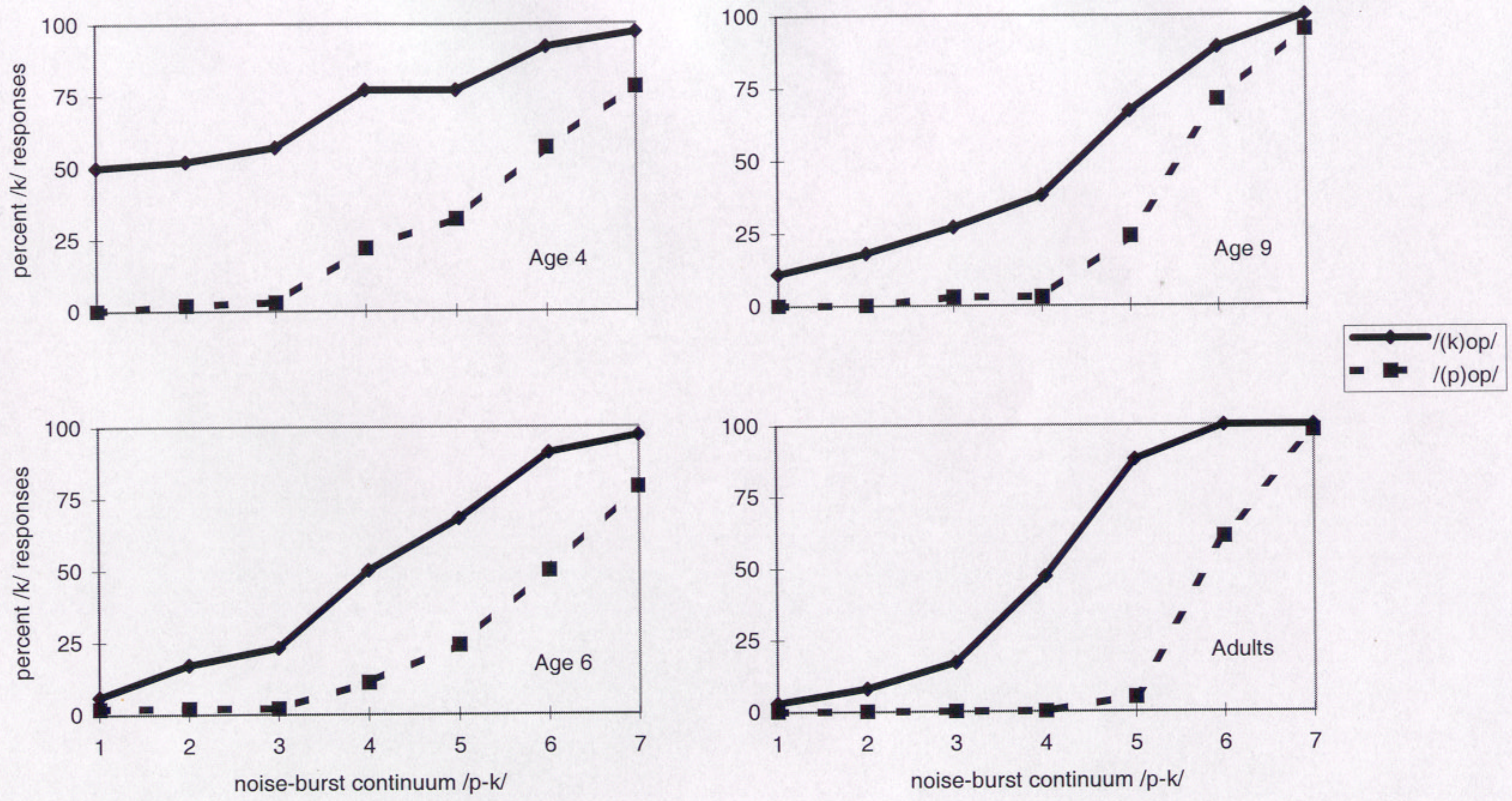


Figure 6.10. Mean proportions resopnses for /pɔp-kɔp/stimuli per age group.

The post-hoc Tukey HSD test with Transition showed that for all groups mean slope for the continuum with /κ(Oπ)/ transitions was significantly lower than that for /p(Oπ)/. Thus, the classification functions for the /p-k/ continuum followed by the vowel from /κOπ/ were shallower than those for /p-k/ continuum followed by the vowel from /πOπ/. This indicates that, according to the listeners, the first continuum contained more ambiguous stimuli than the second continuum.

The results of a post-hoc Tukey HSD test with factor Age showed that there were significant differences between the mean slope values of each age group. The mean values for the 4-year-olds are lowest, and the slope value increases with age. Even the function gradients of the 9-year-olds are shallower than those of the adult listeners. This means that the weighting of the noise burst increases as children grow older, but is still not fully developed at age 9.

The separation between the functions was determined to see the influence of the formant transition cue on the labelling responses. The effects of the factors Age and Stimulus were computed with a two-way analysis of variance on the groups' arcsine-difference data. Age had a significantly effect at F=7.27 and p<.001, and Stimulus had a significant effect at F=9.01, p<.001. In addition, there was a significant Age by Stimulus interaction, F=2.152, p=.005. The post-hoc Tukey HSD test with Age revealed that the youngest group of children demonstrated a significantly larger shift size than the other age groups. This means that the 4-year-olds weighted the formant-transition cue relatively more heavily than the older children and adults did.

In table 6.2 it can be seen that the shift sizes for the 6-year-olds' and 9-year-olds' functions appear smaller than the shift size of the adults' functions, but this difference is not significant. There are clear differences in shift size between groups as a function of the stimulus. In figure 6.10 it can be seen that adults adjust their weighting to the strength of the cue: the formant transition cue is weighted most strongly with the ambiguous noise portions. In the graphs with the results of the 6- and 9-year-olds it can be seen that the distance between their functions does not change much as an effect of the ambiguity of the noise cue. The post-hoc Tukey HSD test for Stimulus revealed that the shift size was largest for stimuli 4, 5 and 6, which indicates that, according to the listeners, these stimuli contained the most ambiguous noise portions.

The analysis of the stop-consonant data leads to the conclusion that only the youngest children, the 4-year-olds, show more weighting of the formant transition cue than the older age groups. All listeners use both cues in reaching a decision about the stop-consonant category. The mean slope values show age-related differences for each age group. This indicates that there is an overall tendency for children's classification of stop consonants to weight the noise cue less heavily than the adults. The weighting of the noise cue is not adult-like yet at age 9, the oldest children in the present study.

6.6.4 Vowel results

Figure 6.11 displays the mean proportions of /ζα κ/ - /ζAκ/ responses for each age group. Only 10 four-year-olds completed the whole test. The data of the 4-year-olds consist of 60 judgements per stimulus and there are 66 judgements for the other age groups. In tables 1 and 2 of Appendix D, the mean proportions of the two functions are listed, along with the standard deviations.

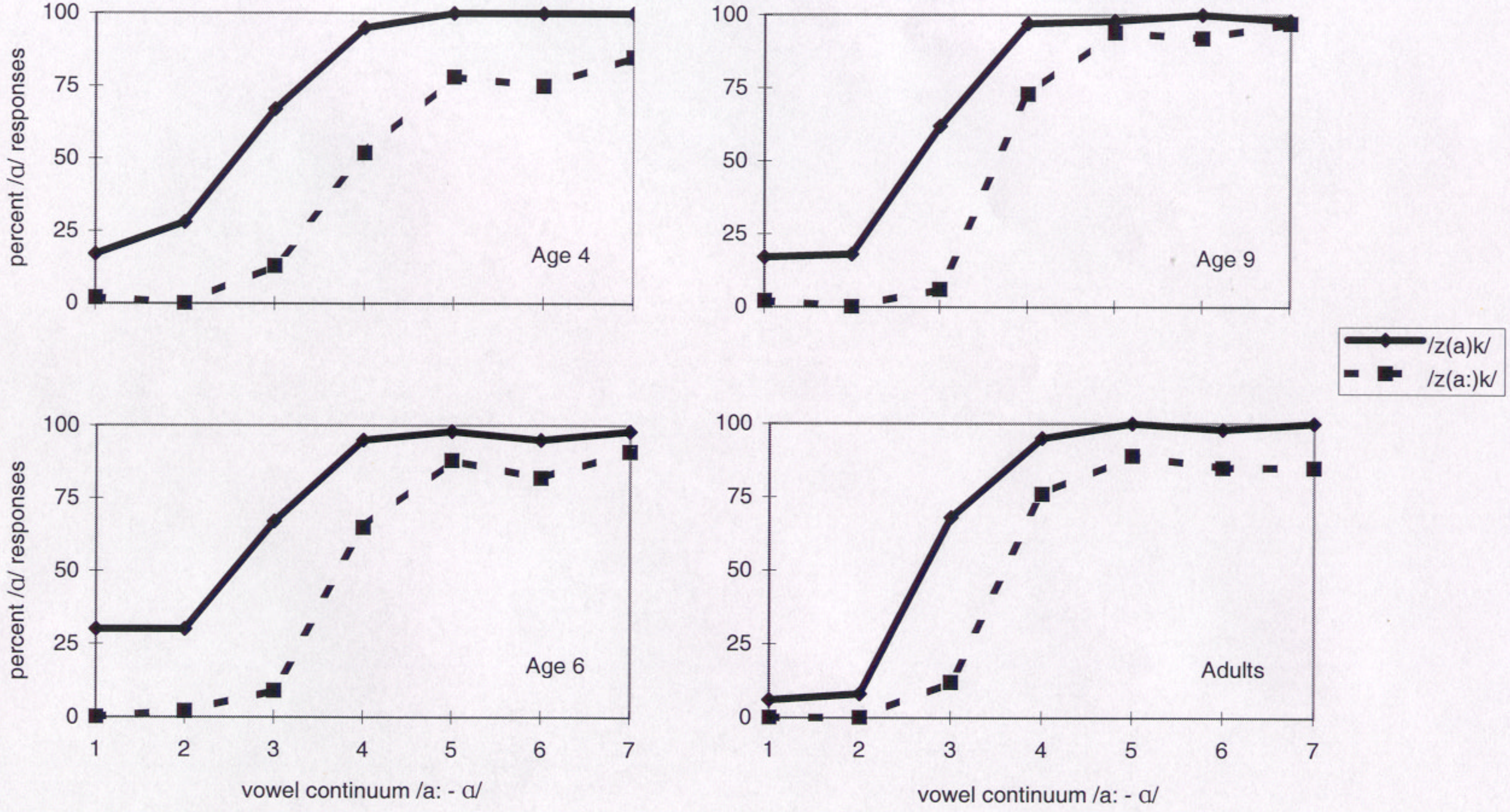


Figure 6.11. Mean proportions responses for /za:k-zɑk/ stimuli per age group.

The figure shows that, even for the adult listeners, only the best exemplars reached asymptotic performance at 0% and 100%. The left endpoint in the continuum with the /ζAκ/ duration reached asymptote at 6%, and the right endpoint in the continuum with the /ζα κ/ duration reached asymptote at 85%. This means that the duration cue was a relatively strong cue compared to the formant-frequency cue. In section 6.5.1.3, in which the vowel stimuli are described, it was already noted that vowel duration is a strong cue in signalling the difference between Dutch /A/ and /α /. Because we expected ceiling effects to occur with the original values, we used slightly less extreme vowel durations. The present results show that even with these ambiguous duration values, some listeners show a tendency to base their phonetic decisions entirely on the duration cue.

During the experiment some adult listeners reported that it was difficult not to let their labelling responses be influenced by the variation in the production of /α / and /A/ in everyday speech. They hesitated in their classification of some of the stimuli and informed the experimenter about their doubts by saying that the vowels sounded ‘as if they were spoken by people with different Dutch dialects’ and that they had the feeling that some stimuli could be identified both as a ‘Southern’ /A/ or as a ‘Northern’ /α /. Although the Utrecht dialect contains an /A/ that may sound like /α / to listeners from other Dutch regions, listeners did not report hearing an Utrecht-like pronunciation of the vowels. The children did not report the adult listeners’ kind of labelling problems, most likely because they have less experience with different Dutch dialects.

In the analysis of variance on the /ζα κ/ - /ζAκ/ data, the effects of Age (4), Duration (2) Stimulus within Duration (7), and Test (3) were computed. The effects of the factors Duration and Stimulus within Duration were found to be significant (F=520.66, p<.001 and F=457.87, p<.001). A comparison of the mean scores of Duration showed that listeners gave more “A” responses when the duration was appropriate for /A/ rather than /α /. There were also significant Age x Duration (F=6.77, p<.001) and Age x Stimulus within Duration (F=2.09, p<.001) interactions. The effect of the manipulation of formant frequency and duration on the listeners’ labelling was examined more closely by means of an analysis of the slope of the functions and the shift size.

Table 6.3. Mean slope and shift size of the /ζα κ/ - /ζAκ/ classification functions for each age group.

	Slope	Shift size
Age 4	.96	.67
Age 6	1.08	.60
Age 9	1.73	.43
Adults	1.83	.43

Concerning the slope, a two-way analysis of variance with factors Age and Duration showed a significant effect of Age (F=4.55, p=.005). A post-hoc Tukey HSD test was performed to analyse pairwise differences between the age groups. The results of this test revealed that slope values were significantly lower for the younger children compared to the older children and the adults. In figure 6.11 it can be seen that there is a reciprocal relationship between the weighting of vowel duration and formant frequencies: the function gradients becomes less shallow when the separation between the functions becomes smaller. The slope of the 9-year-olds is adult-like. These findings demonstrate that the 4- and 6-year-olds weighted the formant frequency cue less heavily than the 9-year-olds and adults did.

The shift-size effect was determined with a two-way analysis of variance with factors Age and Stimulus. This analysis revealed significant effects for both factors, Age, F=3.64, p=.013 and Stimulus, F=14.67, p<.001. The effect of Age is rather small and a post-hoc Tukey HSD test with Age did not show any significant subgroups. The mean arcsine-difference values are displayed in Table 6.3. As can be seen in this table, shift size is largest between the functions of the youngest children; this shift is smaller for the older children and adults. Post-hoc Tukey HSD with factor Stimulus revealed that largest shift size between the function was at stimulus 3.

The /ζα κ/ - /ζΑκ/ results show that all listeners use both vowel spectrum and vowel duration to decide about vowel identity. The analysis of the slope of the classification functions shows that the 4-year-olds and 6-year-olds weight the information in the vowels' formant frequencies less heavily than the older children and the adults do. There are no significant differences in the weighting of vowel duration between age groups. Apparently, the weighting of this vowel-specific temporal cue is adult-like at age 4.

6.7 Discussion

In the present study, the perceptual weighting of acoustic cues in children's and adults' classification of fricatives, stop consonants and vowels was investigated. Evidence was found for developmental differences between children's and adults' cue-weighting schemes.

6.7.1. Cue-weighting for consonants

The fricative and stop-consonant results demonstrate that 4-year-old children assigned more weight to formant transitions and less weight to noise spectra than older children and adults did. At age 6, formant-transition weighting was adult-like. However, weighting of the fricative noise was not adult-like until age 9, and none of the children weighted the noise burst in stop consonant perception as heavily as the adults did.

There was a clear difference between listeners' weighting of the formant-transition cue in fricative and stop-consonant perception. In order to decide about the fricative category, 6- and 9-year-old children and adults used only the information in the noise cue. This was not the case for stop-consonant labelling, in which all listeners depended on both the noise and the formant transition cues and even the adult listeners showed a large effect of the transition cue. The difference may be due to two differences between the acoustic cues for these two consonant contrasts. Firstly, the noise burst was much shorter than the fricative noise (30 ms and 190 ms, respectively) and hence may have contained relatively less 'phonetic information'. Secondly, the stop-consonant formant-transition cue changed in direction as an effect of the change in place of articulation. It was slightly rising for /πO/, but falling for /κO/. This was not the case for formant transitions in the two fricative continua, which were both falling. Maybe the change in direction of the stop-consonant transitions enhances their usefulness as perceptual cues relative to the unchanging fricative transitions.

The cue-weighting results for the fricatives generally support the findings by Nittrouer and colleagues (e.g. 1992, 1996a, 2000b). Yet, there are some differences. In our /Σ - σ/ experiment, the children's weighting schemes for the transition cue are adult-like at age 6. Furthermore, it was shown that the 6- to 9-year-old children and the adults only used the information in the noise cue (the two functions coincided). These two findings are different from those of Nittrouer since, firstly, in her fricative studies even 7-year-olds gave more weight to formant transition cues than adults. Secondly, in her adult results there is also a marked separation between the functions (representing the different formant transitions), indicating that labelling is based on both the information in the noise spectra and in the formant transitions. This discrepancy may be explained by differences between the stimuli. Firstly, we used natural noises concatenated with natural vocalic portions, whereas in Nittrouer's studies synthetic noises were combined with natural vocalic portions (Nittrouer & Studdert-Kennedy, 1987; Nittrouer, 1992, 1996a; 2000b; Nittrouer & Miller, 1997b). The older children in Nittrouer's studies may have been less sensitive to the noise cue because of its synthetic quality, and may therefore have assigned more weight to the natural formant transition cue. Secondly, the differences between the results may be caused by the vowel-related difference in the size of the transition effect that was mentioned by Nittrouer (1992). In that study it was shown that the effect of the formant transition cue was most evident in the /u/ context as opposed to /i/ and /A/. For /i/, the transition effect was rather small: for 3-year-olds the effect was one-seventh as large as the effect found for /u/. The present /ΣOκ-σOκ/ transition effect is similar to the transition effect (or the lack of such an effect) found by Mayo (1999). Mayo's results with /Σo-σo/ stimuli demonstrate no separation between adults' classification functions, so adult listeners only use the information in the noise spectra. Unfortunately, as in the current study, Mayo (1999) only used one fricative-vowel combination and thus the results cannot be compared across vowel contexts.

Our stop-consonant results confirm the findings by Parnell and Amerman (1978) and Ohde et al. (1995, 1997). Although they used a completely different design, they found the use of formant transition cues for stop-consonant labelling to be adult-like at age 6. This is also true in our study: only 4-year-olds assign more weight to the formant-transition cue; from age 6 on, formant transition weighting is adult-like. Like Parnell and Amerman, we also found differences in the weighting of the noise cue up to age 6 or 9. However, Ohde et al. do not report age-related differences in the relative use of the noise cue. This may be due to their manipulation of this cue: in their 'noise-burst condition', they presented the noise burst plus one voiced period, whereas we manipulated the noise burst only.

The developmental patterns for cue-weighting of consonant contrasts found in this study are in line with findings on classification of consonant contrasts with stimuli in which a single acoustic property was systematically varied along an acoustic continuum between two phonemes (e.g. Burnham, Earnshaw & Clark, 1991; Krause, 1982; Kuijpers, 1996; Sussman & Carney, 1989; Wolf, 1973; Zlatin & Koenigsknecht, 1975). According to the results in those studies, phoneme categorisation behaviour of children as old as 10 to 12 years of age is still less consistent than adults' classification.

6.7.2 Cue-weighting for vowels

The current vowel results demonstrate that all listeners integrated the information in the vowel spectra with vowel duration in order to categorise the vowels. There are age-related differences in weighting of the formant-frequency cue, but there are no age-related differences between the weighting of the duration cue. All listeners based their vowel judgements on both vowel cues. We expected that children's weighting schemes for vowels would be adult-like earlier in development than the schemes for consonants since vowel perception is said to be acquired at an earlier age in general. The present results partly confirm our expectations: children and adults weighted the vowel duration cue in similar ways, but the weighting of the formant frequencies was less for the 4- and 6-year-olds than for the older age groups. This means that children's weighting of the acoustic cues for vowel categorisation is not fully developed at age 6, which is at a much older age than was expected on the basis of findings in previous studies (Vihman, 1996; Werker & Polka, 1993).

6.7.3 The development of the lexicon

In the introduction section of this chapter it was suggested that children's perceptual weighting results indicate that learning must be involved in acquiring language-specific weighting strategies. In the course of speech development children become familiar with the variability in the amount of information conveyed by each acoustic cue across phonetic environments, which guides the developmental shift towards adult-like perception. In other developmental speech perception studies, age-related differences in perceptual strategies are also explained by age-related differences in the size of the lexicon, especially in the number of phonologically similar words (e.g. Charles-Luce & Luce, 1990, 1995; Jusczyck, 1993; Walley, 1993). Studies that describe the size of children's lexicons appear to be very scarce. To date, the only published Dutch study is the one by Kohnstamm, Schaerlaekens, DeVries, Akkerhuis, and Froonincksx (1981). In this study, age-of-acquisition ratings (AoA) were used to determine the size of the passive vocabulary of 6-year-olds. In AoA studies, the age at which a particular word or concept is learned (passive knowledge) is determined by asking participants to estimate the age at which various words are learned. In Kohnstamm et al. (1981), the AoA estimates were obtained by asking teachers to indicate for 6,785 words (two- to eight-letter nouns and verbs) to what extent they thought that each word was known to 6-year-olds. From the total number of selected words, 1,278 words were estimated to be known to 6-year-olds. Unfortunately, the Kohnstamm et al. (1981) ratings are limited, because they only provide information about one moment in time: the transition from kindergarten to primary school.

Recently, a new (Flemish) Dutch AoA rating was performed by De Moor, Ghyselinck, and Brysbaert (in press). In this study, AoA ratings were estimated for 2,816 four- and five-letter nouns from a large corpus of Dutch words (CELEX Database: Baayen, Piepenbrock & Van Rijn, 1993). The rating participants were 559 undergraduate students from the University of Gent (Belgium). This time, the age range was not limited to 6-year-olds (as in Kohnstamm et al., 1981), but the participants were asked to select an age between 2 to 12 years. The ratings show that for 143 words the AoA was estimated at age 4 or younger. This is 5% of the total number of 4- and 5-letter words in the CELEX database. At age 6, word knowledge has increased to 606 words. This is 4.24 times the words learned at age 4, and it is 22% of the total number of selected words. At age 9, children were estimated to know 1,534 words (54% of the selected 2,816 nouns). This is 10.73 times the word knowledge of 4-year-olds, and 2.53 times that of 6-year-olds.

These AoA ratings imply that there is a marked increase in the number of 4- and 5-letter nouns between ages 4 and 6. This growth is almost twice as large as the increase between the ages of 6 and 9. The vocabulary spurt between 4 and 6 years of age may be a consequence of children's entry into school-life. The school environment will provide a totally new input to the child and therefore lead to a significant increase of their lexicon. Another increase of children's vocabulary is expected when they begin to read at approximately age 6

(Adams, 1990). There might be a relationship between the shift in children's weighting of the formant-transition cue and their relatively large vocabulary growth between ages 4 and 6. However, this remains a speculative suggestion, since we did not investigate the size of the children's lexicons. It would be interesting to see if future research in lexical development could be directly related to age-related differences in speech perception.

6.4 Further research

Today, the primary theoretical and empirical focus in speech perception research is on the performance and development of infants. As a result, techniques for studying perception are much better established for infants than for slightly older children. The consequence is that a complete picture of the development of phonetic representation from infancy to early childhood is, for the time being, unobtainable (but see Swingley & Aslin, 2000). In the previous section we have already suggested that it would be interesting to investigate the development of the lexicon in relation to speech perception. In this section some other suggestions are made for further research, specifically concerning the perceptual integration of acoustic speech cues.

6.4.1 The development of phonemic awareness

In Nittrouer (1996b), it is suggested that there is a relation between phonemic awareness and the fricative weighting results. Phonemic awareness[3] refers to the ability to think about and manipulate speech units. By four years of age most children develop an awareness of subsyllabic units such as onset and rime. Only at approximately the age of six, children start learning to identify individual phonemes in words (see overview in Stackhouse & Wells, 1997, and in Bishop, 1997).

In Nittrouer's study (1996b), children (ages 7 and 8) with a low socio-economic status and/or with chronic otitis media and a control group were tested. One of the goals of the study was to determine if developmental changes in weighting strategies were correlated with developmental changes in phonemic awareness. The weighting of fricative noise and vocalic transitions was investigated with the fricative stimuli /su-Σu/ and /sA-ΣA/; the phonemic awareness tasks were phoneme deletion and pig-Latin. The results demonstrated a link between perceptual weighting strategies and phonemic awareness: children who were more advanced developmentally on one task, were generally found to be more advanced on the other task. Unfortunately, nothing can be concluded about the direction of causality between the development of these processes. In Mayo (1999), the relation between perceptual weighting and phonemic awareness was tested with normally developing children, aged 5 to 7. The results demonstrate that high levels of phonemic awareness were related to adult-like perceptual weighting schemes. Furthermore, the results demonstrated that children who had not received literacy or phonemic awareness training, displayed no adult-like weighting

---

[3] Phonemic awareness measures are, for instance, phoneme blending and phoneme segmentation, phoneme deletion and language games, such as pig-Latin. In phoneme-deletion tasks a nonsense word is presented and one phoneme has to be removed in order to make a real word. Pig-Latin words are formed by taking the initial consonant of an English word and moving it to the end of the word, then adding "-ay" after the consonants. Thus, "You must be kidding!" would be fashioned "Ouyay ustmay ebay iddingkay !".

schemes. It is suggested that in the absence of phonemic awareness development, perceptual weighting remains non-adult-like, even in older children (Mayo, 1999). However, this interpretation is only based on data of four children and therefore warrants further investigation.

Although both studies try to link phonemic awareness to perceptual strategies, we do not think it is very likely that phonemic awareness changes word representations in long-term memory. In school, children are given explicit phonemic awareness training as part of learning to read and write. Several phonemic awareness studies have demonstrated that children who are good at these manipulation tasks are also good readers, but phonemic awareness is not a prerequisite for learning to read and write (see review in Bishop, 1997).

As we have said before, it is expected that beginning to read and write leads to a large increase in children's vocabulary (Adams, 1990). Therefore, it may be that the results of Nittrouer (1996b) and Mayo (1999) show the effect of literacy training and the development of the lexicon on children's weighting of acoustic cues and not a unique effect of phoneme awareness.

6.4.2 Language disorders

In Nittrouer (1999), it was shown that, in making fricative labelling decisions, children with poor phonological processing abilities weighted formant transitions more heavily than control children did. According to Nittrouer, this indicates a delay in the developmental weighting shift: the children with poor phonological processing abilities were still using the formant-transition cue in a way shown by younger children, whereas the control children already weighted the noise and formant-transition cues in a way similar to adult listeners.

There are numerous reports that demonstrate that children with Specific Language Impairment or children with dyslexia have impairments in expressive and receptive phonology (see overview Bishop, 1997; and also Leonard, 1998). Of special interest are the claims of Tallal and colleagues (e.g. Tallal & Stark, 1981; Tallal et al., 1996) that refer to the perception of formant-transition information. In Tallal's hypothesis of a "temporal processing deficit" it is said that children with Specific Language Impairment have marked difficulty processing brief and rapidly changing acoustic information, notably formant transitions (but see Mody, Studdert-Kennedy & Brady, 1997). However, the results of Nittrouer (1999) contradict Tallal's hypothesis: children with poor phonological processing abilities based their fricative classification more on formant transitions than the children with good phonological processing abilities did. No evidence was found in favour of a temporal processing deficit. However, the children tested in Nittrouer's study were not severely language- or reading-impaired, as was the case in Tallal's studies. It would be worthwhile to use the cue-weighting paradigm to explicitly test the "temporal processing deficit" hypothesis. If SLI children and children with developmental dyslexia truly have a severe disorder in the processing of formant transition information, this might be reflected in a different weighting of this specific cue by these children compared to normally developing control children.

6.5 Conclusion

In the present study it was shown that young children are able to identify words like “sok-sjok”, “pop-kop”, and “zak-zaak” in a way similar to adults. However, children differ in the relative weights they assign to the acoustic cues that underly the phonetic differences between these minimally differing word pairs. When children grow older and gain more experience with their native language, they adjust their weighting of acoustic speech cues. This results in a shift away from their early weighting schemes towards adult-like cue weighting.

The perceptual cue-weighting findings indicate that children are able to use small acoustic differences between different realisations of speech sounds to decide about the phoneme category they hear. The idea that children’s initial representation of words is not well specified seems to be in contradiction with their ability to focus on such fine-grained phonetic details. The present findings suggest that children’s lexical representations of words may be well specified, as is also suggested in recent studies with much younger children (e.g. Jusczyk & Aslin, 1995; Swingley & Aslin, 2000).

Chapter 7

General discussion and conclusion

The present study deals with the way listeners classify speech sounds that belong to the phonological system of their native language. It was investigated how children learn to assign the appropriate weights to acoustic cues that specify particular speech sounds and how the acquired classification system influences the way adult listeners perceive acoustic differences between speech stimuli.

With respect to the latter question, the hypothesis was tested if there is a strong relationship between listeners' classification and discrimination of stimuli along a continuum between two speech sounds (the 'categorical perception hypothesis'). It was expected that listeners would be relatively poor at detecting differences between within-phoneme-category stimuli, because these stimuli contain acoustic differences that do not convey meaning and thus are irrelevant for speech comprehension. On the other hand, it was thought that listeners would be relatively good at detecting differences between across-phoneme-category stimuli. In Chapter 5, in which a summary and a discussion of the results of this part of our study are presented, we conclude that the categorical perception hypothesis has to be revised: there is no strict relationship between discrimination and classification of speech sounds. There appear to be marked effects of several procedural factors, especially of the discrimination task, on listeners' discrimination performance and hence on the degree of categorical perception. The absence of such a relationship implies that listeners' language use and language development do not have a strong effect on their discrimination of speech. Listeners do not 'lose' their auditory ability to detect non-phonemic acoustic differences between speech sounds.

This conclusion is supported by the findings of cross-linguistic speech perception studies. Initially, there were several reports in which it was claimed that there is a loss of sensitivity to non-native phoneme contrasts as an effect of learning the phonological system of the native language (e.g. Werker et al., 1981). However, what these early reports seem to show is listeners' *classification* rather than their *discrimination* performance. In cross-linguistic studies, usually the change-no change procedure is used, with a non-native phoneme contrast presented in two naturally produced syllables. In our view, even though non-native phonemes are presented, listeners still try to match the stimulus with the most closely resembling native phoneme category. If both stimulus tokens are classified within the same category, change-no change 'discrimination' performance is expected to be very poor, which is what is shown in studies that have used this procedure. But listeners may be able to use two different labels for the non-native tokens, which explains why some non-native phoneme contrasts are 'discriminated' more easily than others (e.g. Best, McRoberts & Sithole, 1988; Werker & Tees, 1984a). The study by Werker and Tees (1984a) is of special interest here since they demonstrated that listeners who could not discriminate non-native contrasts with the change-no change task, readily discriminated the same stimuli when AX discrimination with feedback was used. This suggests that the early studies do not show that there is a loss of discrimination between non-native speech sounds, but that insufficiently sensitive discrimination procedures have been used to reveal listeners' true discrimination performance.

In the developmental part of the present study we focused on speech classification only. It was investigated how children learn the weighting schemes for acoustic speech cues that underlie decisions about phoneme categories. The results show that children integrate cues in a similar way as adult listeners do, but that they assign different weights to these cues. As children grow older and gain experience with the native language, their weighting of cues gradually shifts towards adult-like weighting. The weighting results imply that children are able to change their speech classification according to small changes in acoustic speech cues. This supports our claim that learning the native language does not have consequences for detecting small differences between speech sounds.

In more recent studies with infants, it is generally agreed that infants' ability to detect small acoustic differences between speech sounds does not become worse as an effect of their development of the native classification system. In several reports it has been shown that young infants (1 to 6 months old) are able to discriminate both native and non-native phoneme contrasts within syllables or words. Between 6 months and 12 months of age there appears to be a change in their perception since they become worse at discriminating between non-native speech sounds. At first sight this may seem to contradict our claim that discrimination does not become worse as an effect of language learning. However, this apparent contradiction is again due to confusion of discrimination and classification. The discrimination task used in the infant studies is the same change-no change test we described in our discussion of cross-linguistic research. We have argued that this test measures adult listeners' speech classification instead of their discrimination. Is this also true for infants? We would argue that young infants (1 to 6 months old) attempt to 'categorise' the speech input. By 'categorisation', we mean the general cognitive ability to structure the environmental input into meaningful groups of signals. This differs from 'classification', which in our view refers to the ability to use category representations stored in long-term memory. Young infants have not developed these categories yet, but the behaviour of infants of 12 months of age shows that they have developed some early form of phoneme category representations in long-term memory. Therefore, it would seem that the change in infants' perception between 6 and 12 months reflects their language acquisition and their earliest effort to classify speech stimuli.

In conclusion, it was shown that categorical perception defined as a strict relationship between discrimination and classification of speech sounds does not exist. There are marked effects of several procedural factors, especially of the discrimination task. If a task is used that is sensitive to the acoustic differences between stimuli, there is no relationship between discrimination and classification results. This means that the acquired classification system does not have a negative effect on the discrimination of speech sounds. The auditory system remains capable of perceiving within-phoneme-category differences. The cue-weighting experiment has revealed that the classification system is not fully developed at age nine. In addition, children's adjustment of the weighting of certain speech cues supports the idea that we remain sensitive to subtle acoustic differences.

The development of the native classification system influences the way children and adult listeners classify speech, but does not diminish their ability to perceive acoustic differences between speech sounds. On the contrary, the classification mechanism actually enhances discrimination, especially of meaningful speech sounds.

References

Abramson, A.S. and Lisker, L. (1970). Discriminability along the voicing continuum; Cross-language tests. *Proceedings of the sixth International Conference of Phonetic Sciences in Prague,* 569-573.

Adams, M. (1990). *Beginning to read.* Cambridge: MIT Press.

Ades, A.E. (1977). Vowels, consonants, speech, and nonspeech. *Psychological Review*, 84, 524-530.

Allen, P. and Wightman, F. (1992). Spectral pattern discrimination by children. *Journal of Speech and Hearing Research*, 35, 222-233.

Anderson, N.H. (1982). *Methods of information integration theory*. New York: Academic Press.

Aslin, R.N., Jusczyk, P.W. and Pisoni, D.B. (1998). Speech and auditory processing during infancy: constraints on and precursors to language. In: D. Kuhn and R. Siegles (Eds.), *A handbook of child psychology: cognition, perception, and language,* New York: Wiley, 147-254.

Baayen, R.H., Piepenbrock, R. and Van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Barton, D. (1980). Phonemic perception in children. In: G.H. Yeni-Komshian, J.F. Kavanagh and C.A. Ferguson (Eds.). *Child Phonology, Volume 2: Perception*. New York: Academic Press.

Beers, M. (1995). *The phonology of normally developing and language-impaired children*. Doctoral Dissertation, University of Amsterdam, the Netherlands. Dordrecht: ICG-printing.

Best, C.T., Morrongiello, B. and Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 29, 191-211.

Best, C.T., Studdert-Kennedy, M., Manuel, S. and Rubin-Spitz, J. (1989). Discovering phonetic coherence in acoustic patterns. *Perception & Psychophysics*, 45, 237-250.

Bishop, D.V.M. (1997). *Uncommon understanding. Development and disorders of language comprehension in children*. Hove: Psychology Press.

Burnham, D.K., Earnshaw, L.J. and Clark, J.E. (1991). Development of Native and Non-Native Bilabial Stops: Infants, Children and Adults. *Journal of Child Language,* 18, 231-260.

Carney, A.E., Widin, G.P. and Viemeister, N.F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 62, 961-970.

Charles-Luce, J. and Luce, P. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205-215.

Charles-Luce, J. and Luce, P. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, 22, 727-735.

Clement, C. J. and Wijnen, F. (1994). Acquisition of vowel contrasts in Dutch. *Journal of Speech and Hearing Research*, 37, 83-89.

Cowan, N. and Morse, P.A. (1986). The Use of Auditory and Phonetic Memory in Vowel Discrimination. *Journal of the Acoustical Society of America,* 79, 500-507.

Creelman, C.D. and Macmillan, N.A. (1979). Auditory phase and frequency discrimination: a comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance,* 5, 146-156.

Crowder, R.G. (1982). Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Human Learning and Memory*, 8, 153-162.

Crowther, C.S. and Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing. *Journal of the Acoustical Society of America*, 92, 711-722.

Crul, Th.A.M. and Peters, H.F.M. (1976). *Auditieve discriminatie test (ADIT).* Amsterdam: Swets and Zeitlinger B.V.

Cutting, J.E. (1982). Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics*, 31, 462-476.

Cutting, J.E., Rosner, B.S. and Foard, C.F. (1976). Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 28, 361-378.

Delattre, P.C., Liberman, A.M., Cooper, F.S. and Gerstman, L.J. (1952). An experimental study of the acoustic determinants of vowel colour: observations of one- and two-formant vowels synthesised from spectrographic patterns. *Word,* 8, 195-210.

Dorman, M.F., Kewley-Port, D., Brady, S. and Turvey, M.T. (1977). Vowel recognition: inferences from studies of forward and backward masking. *Quarterly Journal of Experimental Psychology*, 29, 483-497.

Dorman, M.F. Studdert-Kennedy, M. Raphael, L.J. (1977) Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues. *Perception & Psychophysics*, 22, 109-122.

Durlach, N.I. and Braida, L.D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, 46, 372-383.

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

Eimas, P.D., Siqueland, E.R., Jusczyk, P. and Vigorito, J. (1971). Speech perception in infants. *Science,* 171, 303-306.

Elfenbein, J.L. and Davis, J.M. (1993). Developmental Patterns of Duration Discrimination. *Journal of Speech and Hearing Research*, 36, 842-849.

Escudero, P. R. (2000). *The perception of English vowel contrasts: acoustic cue reliance in the development of L2 perception.* Paper presented at the Fourth International Symposium on the Acquisition of Second Language Speech, New Sounds 2000, Amsterdam, The Netherlands.

Foard, C.F. and Kemler-Nelson, D.G. (1984). Holistic and analytic modes of processing: the multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General,* 113, 94-111.

Fourcin, A.J. (1978). Acoustic patterns and speech acquisition. In: N. Waterson and C. Snow (Eds.), *The development of Communication,* London: John Wiley and Sons, 47-72.

Fry, D.B., Abramson, A.S., Eimas, P.D. and Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5, 171-189.

Fujisaki, H. and Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 29, 207-214.

Fujisaki, H. and Kawashima, T. (1971). A model of the mechanisms for speech perception: Quantative analyses of categorical effects in discrimination. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 30, 59-68.

Gerken, L.A. (1992). Child phonology. In: M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, New York: Academic Press, 781-820.

Gerrits, E. and Schouten, M.E.H. (1998). Categorical Perception of Vowels. *Proceedings of the 5th International Conference on Spoken Language Processing in Sydney,* Vol. 6, 2279-2283.

Greenlee, M. (1980). Learning the phonetic cues to the voiced-voiceless distinction: a comparison of child and adult speech perception. *Journal of Child Language*, 7, 459-468.

Groenen, P.A.P. (1997). *Central auditory processing disorders: a psycholinguistic approach.* Doctoral Dissertation, Nijmegen University, The Netherlands.

Hall, J.W. and Grose, J.H. (1991). Notched-noise measures of frequency selectivity in adults and children using fixed-masker-level and fixed-signal-level presentation. *Journal of Speech and Hearing Research*, 34, 651-660.

Hanson, V.L. (1977). Within-category discrimination in speech perception. *Perception and Pychophysics,* 21, 423-430.

Harnad, S. (Ed.) (1987). *Categorical Perception.* New York: Cambridge University Press.

Hazan, V. and Barrett, S. (1999). The development of phoneme categorisation in children aged 6 to 12. *Proceedings of the International Conference of Phonetic Sciences in San Francisco*, 2493-2496.

Heller, L.M. and Trahiotis, C. (1995). The discrimination of samples of noise in monotonic, diotic and dichotic conditions. *Journal of the Acoustical Society of America*, 97, 3775-3781.

Hessen, A.J. van (1988). A new speech modulation model based on sinusoidal representation. *Progress Report Institute of Phonetics Utrecht*, 13, 16-28.

Hessen, A.J. van (1991). *Discrimination of familiar and unfamiliar speech sounds.* Doctoral dissertation, Utrecht University, The Netherlands.

Hessen, A.J. van and Schouten, M.E.H. (1992). Modelling phoneme perception. II: A model of stop consonant discrimination. *Journal of the Acoustical Society of America,* 92, 1856-1868.

Hessen, A.J. van and Schouten, M.E.H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, 56, 56-72.

Heuven, V.J. van, Van Houten, J.E. and De Vries, J.W. (1986). De perceptie van Nederlandse klinkers door Turken. *Spektator*, 15, 225-238.

Howell, P. and Rosen, S. (1984). Natural auditory sensitivities as universal determiners of phonemic contrasts. *Linguistics,* 21, 205-235.

Jusczyk, P.W. (1992) Developing phonological categories from the speech signal. In: C.A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological development: Models, Research, Implications*, Parkton MD: York Press, 17-64.

Jusczyk, P.W. (1993). From general to language-specific capabilities; the WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3-28.

Jusczyk, P.W. and Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology,* 29, 1-23.

Jusczyk, P.W. and Bertoncini, J. (1988). Viewing the development of speech perception as an innately guided learning process. *Language and Speech*, 31, 217-238.

Kaplan, H.L., Macmillan, N.A. and Creelman, C.D. (1978). Methods & Designs. *Behavior Research Methods & Instrumentation*, 10, 796-813.

Kirk, R.E. (1982). *Experimental design. Procedures for the behavioral sciences*. Belmont CA: Brooks/Cole Publishing Company.

Kluender, K.R. (1994). Speech perception as a tractable problem in cognitive science. In: M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, New York: Academic Press, 173-217.

Kluender, K.R., Diehl, R. and Wright, B. (1988). Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics*, 16, 153-169.

Kluender, K.R. and Lotto, A.J. (1994). Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. *Journal of the Acoustical Society of America*, 95, 1044-1052.

Kohnstamm, G.A., Schaerlaekens, A.M., De Vries, A.K., Akkerhuis, G.W. and Froonincksx, M. (1981). *Nieuwe streeflijst woordenschat*. Lisse: Swets & Zeitlinger.

Koopmans-Van Beinum, F.J. (1973). Comparative phonetic vowel analysis. *Journal of phonetics*, 1, 249-261.

Kouwenaar, G. (1971). '*gedacht*' uit Data/decors. Querido. Reproduced with permission.

Krause, S.E. (1982). Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *Journal of the Acoustical Society of America*, 71, 990-995.

Kuhl, P.K. (1981). Discrimination of speech by non-human animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America,* 70, 340-349.

Kuhl, P.K. (1985). Methods in the study of infant speech perception. In: G. Gottlieb and N.A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life,* Norwood NJ: Ablex, 223-251.

Kuhl, P.K. and Padden, D.M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, 32, 542-550.

Kuijpers, C.T.L. (1991). Temporal control in speech of children and adults. *Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en-Provence: Université de Provence*, 4, 150-153.

Kuijpers, C.T.L. (1996). Perception of Voicing Contrast by Dutch Children and Adults. *Journal of Phonetics,* 24, 367-382.

Kunisaki, O. and Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 11, 85-91.

Ladefoged, P. (1967). *Three areas of experimental phonetics.* London: Oxford U.P.

Leonard, L.B. (1998). *Children with specific language impairment.* Cambridge: MIT Press.

Liberman, A.M. (1996). *Speech: A Special Code*. Cambridge, MA: MIT Press.

Liberman, A.M., Cooper, F.S., Shankweiler, D.S. and Studdert-Kennedy, M. (1967). Perception of the Speech Code. *Psychological Review*, 74, 431-461.

Liberman, A.M., Harris, K.S., Kinney, J.A. and Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61, 379-388.

Liberman, A.M., Mattingly, I.G. and Turvey, M. (1972). Language codes and memory codes. In: A.W. Melton and E. Martin (Eds.), *Coding processes and human memory*, Washington DC: Winston, 307-334.

Liberman, I.Y., Shankweiler, D., Fischer, F.W. and Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental child Psychology*, 18, 201-212.

Mackain, K.S., Best, C.T. and Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369-390.

Macken, M.A. (1992). Where's phonology? In: C.A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological development: Models, Research, Implications,* Parkton MD: York Press, 249-272.

Macmillan, N.A. and Creelman, C.D. (1991). *Detection Theory: A User's Guide*. New York: Cambrigde University Press.

Macmillan, N.A., Kaplan, H.L. and Creelman, C.D. (1977). The psychophysics of categorical perception. *Psychological Review,* 84, 452-471.

Macmillan, N.A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In: S. Harnad (Ed.), *Categorical Perception,* Cambridge/ New York. 53-85.

Macmillan, N.A., Goldberg, R.F. and Braida, L.D. (1988). Resolution for Speech Sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*, 84, 1262-1280.

Mann, V.A. and Repp, B.H. (1980). Influence of vocalic context on the perception of the /Σ/-/σ/ distinction. *Perception & Psychophysics*, 28, 213-264.

Mann, V.A., Sharlin, H.M. and Dorman, M. (1985). Children's perception of sibilants: The relation between articulation and perceptual development. *Journal of Experimental Child Psychology,* 39, 252-264.

Markel, J.D. and Gray, A.H. (1976). *Linear prediction of speech.* New York: Springer-Verlag.

Massaro, D.W. (1972a). Perceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 79, 124-145.

Massaro, D.W. (1972b). Stimulus information vs processing time in auditory pattern recognition. *Psychonomic Society*, 12, 50-56.

Massaro, D.W. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology*, 102, 199-208.

Massaro, D.W. (Ed.) (1987). *Speech Perception by Ear and Eye.* London: Lawrence Erlbaum Associates.

Massaro, D.W. (1994). Psychological aspects of speech perception: implication for research and theory. In: M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, New York: Academic, 219-263.

Massaro, D.W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle.* Cambridge, Mass.: MIT Press.

Massaro, D.W. and Cohen, M.M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2, 15-35.

Mayo, C. (1999). Perceptual weighting and phonemic awareness in pre-reading and early-reading children. *Proceedings of the International Conference of Phonetic Sciences in San Francisco*, 175-178.

Metsala, J.L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*, 25(1), 47-56.

Mody, M., Studdert-Kennedy, M. and Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, 64, 199-231.

Moor, W. de, Ghyselinck, M. and Brysbaert, M. (in press). Verwervingsleeftijden van 2816 Nederlandse vier- en vijf letterwoorden. *Psychologica Belgica*.

Moore, B.C.J. (1997). *An introduction to the psychology of hearing*. San Diego: Academic Press.

Morosan, D.E. and Jamieson, D.G. (1989). Evaluation of a technique for training new speech contrasts: generalisation across voices, but not word position or task. *Journal of Speech and Hearing Research,* 26, 501-511.

Morrongiello, B.A., Robson, R.C., Best, C.T. and Clifton, R.K. (1984). Trading relations in the perception of speech by 5-year old children. *Journal of Experimental Child Psychology,* 37, 231-250.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20, 351-382.

Nittrouer, S. (1996a). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech and Hearing Research,* 39, 278-297.

Nittrouer, S. (1996b). The relation between speech perception and phonemic awareness: evidence from low-SES children and children with chronic OM. *Journal of Speech and Hearing Research,* 39, 1059-1070.

Nittrouer, S. (1999). Do temporal processing deficits cause phonological processing problems? *Journal of Speech, Language and Hearing Research*, 42, 925-942.

Nittrouer, S. (2000a). *Learning to apprehend phonetic structure from the speech signal: the hows and whys*. Paper presented at the VIIIth Meeting of the International Clinical Phonetics and Linguistics Association in Edinburgh, 16-18 August 2000.

Nittrouer, S. (2000b). The effect of segmental order on fricative labelling by children and adults. *Perception & Psychophysics*, 62, 266-284.

Nittrouer, S. and Crowther, C.S. (1998). Examining the role of auditory sensitivity in the developmental weighting shift. *Journal of Speech, Language and Hearing Research*, 41, 809-818.

Nittrouer, S., Crowther, C.S. and Miller, M.E. (1998). The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults. *Perception & Psychophysics*, 60, 51-64.

Nittrouer, S., Manning, C. and Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of the Acoustical Society of America*, 94, S1865.

Nittrouer, S. and Miller, M.E. (1997a). Developmental weighting shifts for noise components of fricative syllables. *Journal of the Acoustical Society of America*, 102, 572-580.

Nittrouer, S. and Miller, M.E. (1997b). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, 101, 2253-2265.

Nittrouer, S. and Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, 30, 319-329.

Nooteboom, S.G. (1972). *Production and perception of vowel duration*. Doctoral dissertation, University of Utrecht, The Netherlands.

Nooteboom, S.G. and Cohen, A. (1984). *Spreken en verstaan: een nieuwe inleiding tot de experimentele fonetiek*. Assen: Van Gorcum.

Ohde, R.N. and Haley, K.L. (1997). Stop consonant and vowel perception in 3- and 4-year-old children. *Journal of the Acoustical Society of America,* 102, 3711-3722.

Ohde, R.N., Haley, K.L. and McMahon, C.W. (1996). A developmental study of vowel perception from brief synthetic consonant-vowel syllables. *Journal of the Acoustical Society of America*, 100, 3813-3824.

Ohde, R.N., Haley, K.L., Volperian, H.K. and McMahon, C.W. (1995). A developmental study of the perception of onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, 97, 3800-3812.

Parnell, M.M. and Amerman, J.D. (1978). Maturational influences on perception of coarticulatory effects. *Journal of Speech and Hearing Research,* 21, 682-701.

Pastore, R.E. (1981). Possible psychoacoustic factors in speech perception. In: P.D. Eimas and J.L. Miller (Eds.), *Perspectives in the study of speech*, Hillsdale: Erlbaum, 165-205.

Pastore, R.E. (1987). Categorical perception: Some psychophysical models. In: S. Harnad (Ed.), *Categorical Perception,* New York: Cambridge University Press, 29-52.

Pastore, R.E., Ahroon, W.A., Baffuto, J., Friedman, C., Puleo, J.S. and Fink, E.A. (1977). Common factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance,* 3, 686-696.

Pastore, R.E., Friedman, C.J. and Buffato, K.J. (1976). A comparative evaluation of the AX and two ABX procedures. *Paper presented at the 92nd meeting of the Acoustical Society of America, San Diego.*

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics,* 13, 253-260.

Pisoni, D.B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition,* 3, 7-18.

Pisoni, D.B. (1985). Speech perception: some new directions in research and theory. *Journal of the Acoustical Society of America*, 78, 381-388.

Pisoni, D.B., Aslin, R.N., Perey, A.J. and Hennesy, B.L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 297-314.

Pisoni, D.B. and Glanzman, D.L. (1974). Decision processes in speech discrimination as revealed by confidence ratings. *Research on speech perception (Progress report no. 1).* Bloomington: Indiana University, 114-128.

Pisoni D.B. and Luce, P.A (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25, 21-52.

Pisoni, D.B. and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15, 285-290.

Pitt, M.A. and Samuel, A.G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance,* 19, 699-725.

Plomp, R. (1964). The rate of decay of auditory sensation. *Journal of the Acoustical Society of America,* 36, 277-282.

Pols, L.C.W. (1979). Coarticulation and the identification of initial and final plosives. In: J.J. Wolff and D.H. Klatt (Eds.), *ASA*50 Speech Communication Papers*, 459-562.

Pols, L.C.W., Tromp, H.R.C. and Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America*, 51, 1093-1101.

Remez, R.E. (1989). When the objects of perception are spoken. *Ecological Psychology*, 2, 161-180.

Repp, B.H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics*, 30, 217-227.

Repp, B.H. (1984). Categorical Perception: Issues, methods, findings. In: Lass, N.J. (Ed.), *Speech and language: Advances in Basic Research and Practice*, 10, 244-322.

Repp, B.H., Healy, A.F. and Crowder, R.G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance,* 5, 129-145.

Rietveld, A.C.M. and Van Heuven, V.J. (1997). *Algemene fonetiek*. Bussum: Uitgeverij Countinho.

Rosen, S.M. and Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception & Psychophysics*, 30, 156-168.

Rosen, S.M. and Howell, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In: S. Harnad (Ed.), *Categorical Perception*, Cambridge/ New York. 113-160.

Sachs, R.M. (1969). Vowel identification and discrimination in isolation vs. word context. *Quarterly progress report (No. 93).* Cambridge, Mass.: MIT, Research Laboratory of Electronics, 169, 220-229.

Samuel, A.G. (1977). The effect of discrimination training on speech perception: noncategorical perception. *Perception & Psychophysics*, 20, 61-65.

Sawusch, J.R., Nusbaum, H.C. and Schwab, E.C. (1980). Contextual effects in vowel perception II: Evidence for two processing mechanisms. *Perception & Psychophysics*, 27, 421-434.

Schouten, M.E.H. (1980). The case against a speech mode of perception. *Acta Psychologica,* 44, 71-98.

Schouten, M.E.H. (Ed.) (1987). *The Psychophysics of Speech perception.* Dordrecht: Martinus Nijhoff Publishers.

Schouten, M.E.H. and Pols, L.C.W. (1979). Vowel segments in consonantal contexts: a spectral study of coarticulation-part I. *Journal of Phonetics*, 7, 1-23.

Schouten, M.E.H. and Pols, L.C.W. (1983). Perception of plosive consonants: The relative contributions of bursts and vocalic transitions. In: M. Van Den Broecke, V. Van Heuven and W. Zonneveld (Eds.) *Sound Structures*. Dordrecht: Foris Publications. 227-243.

Schouten, M.E.H. and Van Hessen, A.J. (1992). Modelling Phoneme Perception, I: Categorical Perception. *Journal of the Acoustical Society of America,* 92, 1841-1855.

Stackhouse, J. and Wells, B. (1997). *Children's speech and literacy difficulties: A psycholinguistic framework*. Gateshead: Athen≇um Press Ltd.

Smits, R., Ten Bosch, L. and Collier, R. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. *Journal of the Acoustical Society of America,* 100, 3852-3864.

Son, R.J.J.H. van and Pols, L.C.W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 88, 1683-1693.

Stevens, K.N. (1981). Constrainst imposed by the auditory system on the properties used to classify speech sounds; Evidence from phonology, acoustics and psychoacoustics. In: T. Myers, J. Laver and J. Anderson (Eds.), *Advances in psychology: the cognitive representation of speech,* Amsterdam: North-Holland. 61-74.

Stevens, K.N., Liberman, A.M., Öhman, S.E.G. and Studdert-Kennedy, M. (1969). Crosslanguage study of vowel perception. *Language and Speech*, 12, 1-23.

Stoep, J. and Verhoeven, L. (in press). Family and classroom predictors of children's early language and literacy development. National Reading Conference Yearbook, 49.

Strange, W. and Jenkins, J.J. (1978). The role of linguistic experience in the perception of speech. In: R.D. Walk and H.L. Pick (Eds.), *Perception and Experience*, New York: Plenum, 125-169.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuo-motor structure. In: A. Allport, D. MacKay, W. Prinz & E. Scheerer (Eds.), *Language perception and production,* London: Academic Press, 67-84.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. and Cooper, F.S. (1970). The motor theory of speech perception: a reply to Lane's critical review. *Psychological Review,* 77, 234-249.

Studebaker, G.A. (1985). A “rationalized’ arcsine transform. *Journal of Speech and Hearing Research,* 28, 474-494.

Sussman, J.E. (1993). Auditory processing in children’s speech perception: Results of selective adaptation and discrimination tasks. *Journal of Speech and Hearing Research*, 36, 380-395.

Sussman, J.E. and Carney, A.E. (1989). Effects of transition length on the perception of stop consonants by children and adults. *Journal of Speech and Hearing Research*, 32, 151-160.

Svachkin, N.K. (1973). The development of phonemic speech perception in early childhood. In: C.A. Ferguson and D.I. Slobin (Eds.), *Studies of child language development,* New York: Holt, Rinehart and Winston, (Original work published in 1948), 91-127.

Swingley, D. and Aslin, R.N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147-166.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Najarajan, S.S., Schreiner, C., Jenkins, W.M. and Merzenich, M.M. (1996). Language comprehension in language impaired children improved with acoustically modified speech. *Science,* 271, 81-84.

Tallal, P. and Stark, R.E. (1981). Speech acoustic cue discrimination abilities of normally developing and language-impaired children. *Journal of the Acoustical Society of America,* 69, 568-574.

Tartter, V.C. (1981). A comparison of the identification and discrimination of synthetic vowel and stop-consonant stimuli with various acoustic properties. *Journal of Phonetics*, 9, 477-486.

Tartter, V.C. (1982). Vowel and consonant manipulations and the dual-coding model of auditory storage: A Re-evaluation. *Journal of Phonetics*, 10, 217-223.

Trahiotis, C. and Bernstein, L.R. (1990). Detectability of interaural delays over select spectral regions: Effects of flanking noise. *Journal of the Acoustical Society of America*, 87, 810-813.

Treiman, R. and Breaux, A.M. (1982). Common phoneme and overall similarity relations among spoken word syllables; their use by children and adults. *Journal of Psycholinguistic Research,* 11, 581-610.

Verbunt, J. (1996). *Temporal integration of the auditory system*. Report no. 1104, Institute for Perception Research, Eindhoven, The Netherlands.

Vihman, M.M. (1996). *Phonological development: The origins of language in the child.* Cambridge/Oxford: Blackwell Publishers.

Waldrip-Fruin, C. and Peach, S. (1984). Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop-consonants. *Language and Speech*, 27, 367-379.

Walley, A.C. (1993). The role of vocabulary development in children’s spoken word recognition and segmentation ability. *Developmental Review*, 13, 286-350.

Walley, A.C. and Carrell, T.D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 3, 1011-1022.

Walley, A.C., Smith, L.B. and Jusczyk, P.W. (1986). The role of phonemes and syllables in the perceived similarity of speech sounds for children. *Memory and Cognition*. 14, 220-229.

Waterson, N. (1978). *Prosodic phonology: The theory and its application to language acquisition and speech processing*. Newcastle upon Tyne: Grevatt & Grevatt.

Werker, J.F., Gilbert, J.H.V., Humphrey, K. and Tees, R.C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-353.

Werker, J.F. and Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37, 35-44.

Werker, J.F. and Pegg, J.E. (1992). Speech perception and phonological acquisition. In: C.A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological development: Models, Research, Implications*, Parkton MD: York Press, 285-311.

Werker, J.F. and Polka, L. (1993). Developmental changes in speech perception: new challenges and new directions. *Journal of Phonetics,* 21, 83-101.

Werker, J.F. and Tees, R.C. (1984a). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.

Werker, J.F. and Tees, R.C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866-1878.

Whalen, D.H. (1981). Effects of vocalic transitions and vowel quality on the English /σ/-/Σ/ boundary. *Journal of the Acoustical Society of America*, 69, 275-282.

Wieringen, A. van (1995). *Perceiving dynamic speechlike sounds*. Doctoral dissertation, University of Amsterdam, The Netherlands.

Wolf, C.G. (1973). The perception of stop consonants by children. *Journal of Experimental Child Psychology,* 16, 318-331.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, 60, 1381-1389.

Zlatin, M.A. and Koenigsknecht, R.A. (1975). Development of the voicing contrast: perception of stop consonants. *Journal of Speech and Hearing Research*, 18, 541-553.

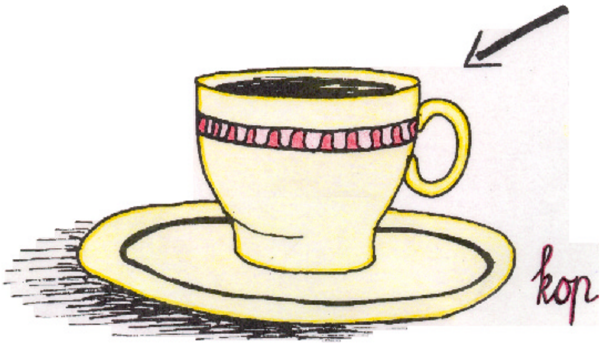
Appendix A

These pictures[*] were used for the experiments with the children. The actual pictures were approximately 25 cm x 20 cm and brightly coloured. Pictures 1 and 2 represent ‘plod’ and ‘sock’, numbers 3 and 4 are ‘doll’ and ‘cup’, and 5 and 6 are ‘bag’ and ‘shop’.

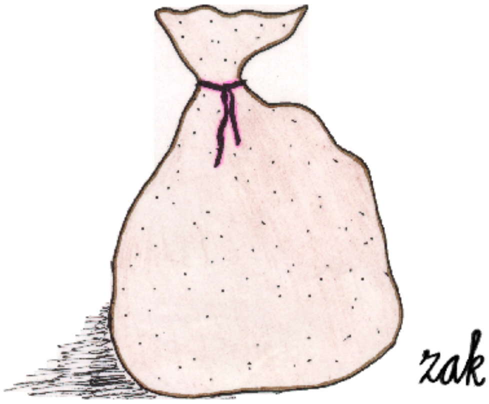
1

2

3



4



5

6

---

[*] With special thanks to Marieke Franken, who made the drawings.

Appendix B

Table 1. Mean proportion and standard deviation per /Σ (Ok)/-transition stimulus and age group.

/Σ/ to /s/ noise		stim1	stim2	stim3	stim4	stim5	stim6	stim7
Age 4	Mean	.03	.02	.03	.27	.84	1.00	.98
	SD	.07	.05	.06	.22	.25	.00	.05
Age 6	Mean	.00	.00	.05	.35	.83	.97	1.00
	SD	.00	.00	.07	.15	.19	.06	.00
Age 9	Mean	.00	.00	.03	.52	.95	1.00	1.00
	SD	.00	.00	.06	.34	.10	.00	.00
Adult	Mean	.00	.00	.00	.35	.89	1.00	1.00
	SD	.00	.00	.00	.31	.19	.00	.00

Table 2. Mean proportion and standard deviation per /σ (Ok)/-transition stimulus and age group.

/Σ/ to /s/ noise		stim1	stim2	stim3	stim4	stim5	stim6	stim7
Age 4	Mean	.03	.08	.23	.71	.86	.97	1.00
	SD	.09	.13	.26	.28	.14	.06	.00
Age 6	Mean	.00	.03	.06	.41	89	.95	1.00
	SD	.00	.06	.08	.19	20	.07	.00
Age 9	Mean	.00	.00	.03	.52	95	1.00	1.00
	SD	.00	.05	.07	.32	10	.00	.00
Adult	Mean	.00	.00	.00	.35	89	1.00	1.00
	SD	.00	.05	.00	.40	10	.00	.00

Appendix C

Table 1. Mean proportion and standard deviation per /pOp/-transition stimulus and age group.

/p/ to /k/ noise		stim1	stim2	stim3	stim4	stim5	stim6	stim7
Age 4	Mean	.00	.02	.03	.22	.32	.57	.78
	SD	.00	.05	.07	.16	.25	.42	.29
Age 6	Mean	.02	.02	.02	.11	.24	.50	.79
	SD	.05	.05	.05	.11	.28	.35	.27
Age 9	Mean	.00	.00	.03	.03	.24	.71	.95
	SD	.00	.00	.06	.06	.21	.26	.14
Adult	Mean	.00	.00	.00	.00	.05	.61	.98
	SD	.00	.00	.00	.00	.07	.30	.05

Table 2. Mean proportion and standard deviation per /k (Op)/-transition stimulus and age group.

/p/ to /k/ noise		stim1	stim2	stim3	stim4	stim5	stim6	stim7
Age 4	Mean	.50	.52	.57	.77	.77	.92	.97
	SD	.32	.46	.40	.31	.26	.15	.10
Age 6	Mean	.06	.17	.23	.50	.68	.91	.97
	SD	.11	.22	.20	.39	.25	.10	.09
Age 9	Mean	.11	.18	.27	.38	.67	.89	1.00
	SD	.19	.31	.39	.39	.30	.15	.00
Adult	Mean	.03	.08	.17	.47	.88	1.00	1.00
	SD	.06	.15	.28	.22	.14	.00	.00

Appendix D

Table 1. Mean proportion and standard deviation per /zα k/-duration stimulus and age group.

/α / to /A/ spectrum		stim1	stim2	stim3	stim4	stim5	stim6	stim7
Age 4	Mean	.02	.00	.13	.52	.78	.75	.85
	SD	.05	.00	.24	.14	.29	.29	.20
Age 6	Mean	.00	.02	.09	.65	.88	.82	.91
	SD	.00	.00	.11	.21	.16	.15	.13
Age 9	Mean	.02	.00	.06	.73	.94	.92	.97
	SD	.00	.00	.15	.29	.13	.15	.06
Adult	Mean	.00	.00	.12	.76	.89	.85	.85
	SD	.00	.00	.25	.31	.28	.31	.28

Table 2. Mean proportion and standard deviation per /ζΑκ/-duration stimulus and age group.

/α / to /A/ spectrum		stim1	stim2	stim3	stim4	stim5	stim6	stim7
Age 4	Mean	.17	.28	.67	.95	1.00	1.00	1.00
	SD	.26	.29	.24	.11	.00	.00	.00
Age 6	Mean	.30	.30	.67	.95	.98	.95	.98
	SD	.25	.23	.28	.07	.05	.10	.04
Age 9	Mean	.17	.18	.62	.97	.98	1.00	.98
	SD	.21	.29	.31	.06	.05	.00	.05
Adult	Mean	.06	.08	.68	.95	1.00	.98	1.00
	SD	.19	.15	.33	.14	.00	.05	.00

Samenvatting

Aan de verwerking van het akoestische spraaksignaal tot een betekenisvolle boodschap liggen meerdere processen ten grondslag. Een van die processen is de segmentatie van het signaal in afzonderlijke fonemen (klinkers en medeklinkers). Dit is geen triviale vaardigheid, want fonemen worden niet als nette, afgebakende akoestische segmenten weergegeven. Bovendien is er een grote variatie in de akoestische weergave van fonemen, ook van hetzelfde foneem, door verschillen tussen sprekers in toonhoogte, luidheid en spreektempo. Toch hebben we geen moeite met het herkennen van fonemen in het spraaksignaal. Als onderdeel van de taalontwikkeling hebben we geleerd wat de belangrijke, betekenisonderscheidende informatie in het akoestische signaal is, zoals kenmerken die bepaalde fonemen in de moedertaal specificeren. Daarnaast hebben we geleerd om irrelevante, betekenisloze informatie in het spraaksignaal te negeren, zoals verschillende uitspraakvarianten van hetzelfde foneem. Dit proces heet categorische spraakperceptie. Sinds de jaren ’60 is er onderzoek gedaan naar de categorische perceptie van spraak. De bevindingen van dit onderzoek hebben geleid tot de ‘categorische-perceptie-hypothese’. Volgens deze hypothese is er een strikte relatie tussen de discriminatie en classificatie van spraakklanken: luisteraars horen alleen verschillen tussen spraakstimuli die ze hebben ingedeeld in verschillende foneemcategorieën. Dit betekent dat het auditieve systeem wordt ‘aangetast’ als gevolg van de verwerving van het fonologische systeem: luisteraars leren blijkbaar niet alleen bepaalde akoestische informatie te negeren, maar ze worden zelfs slechter in het waarnemen van de niet-betekenisonderscheidende informatie. In de afgelopen decennia is de discriminatie en classificatie van spraakklanken regelmatig onderzocht. Een strikte relatie tussen classificatie en discriminatieresultaten is echter zelden gerapporteerd: de discriminatie is vaak beter dan voorspeld door de classificatie. Bovendien blijkt er een grote variatie te zijn in de relatie tussen de discriminatie- en classificatieresultaten. Dit heeft geleid tot vraagtekens bij de categorische-perceptie-hypothese. Is er wel een robuuste relatie tussen discriminatie en classificatie van fonemen? Factoren die een effect lijken te hebben op deze relatie en dus op de mate van categorische perceptie zijn het type discriminatietaak, de natuurlijkheid van het stimulusmateriaal, de foneemklasse en het interstimulus interval. Het huidige onderzoek had als doel het effect van deze factoren in kaart te brengen. Daarnaast is onderzocht hoe de ontwikkeling van categorische perceptie verloopt.

In hoofdstuk 2 wordt een experiment beschreven waarin het verschil in de mate van categorische perceptie werd onderzocht als een functie van stimulusnatuurlijkheid. Het stimulusmateriaal bestond uit de klinkers /u/ en /i/, uitgesproken in de woorden /pup/ en /pip/. Om een verschil in natuurlijkheid te creëren werden deze woorden uitgesproken in twee uitspraakcondities: in een woordenreeks (woordklinkers) en in een snel voorgelezen tekst (tekstklinkers). De tekstklinkers werden verondersteld meer overeen te komen met klinkers in natuurlijke spraakdialogen dan de klinkers in de geïsoleerde woorden. Er werd verwacht dat de mate van categorische perceptie voor de tekstklinkers groter zou zijn dan voor de woordklinkers.

Voor beide klinkercondities werd een stimuluscontinuüm gegenereerd. Dit gebeurde met behulp van interpolatie van de relatieve amplitude van de spectraal omhullende van de originele klinkers. Op deze manier ontstonden twee reeksen van 8 natuurlijk klinkende stimuli in een continuüm van /pup/ naar /pip/. Deze klinkerreeksen werden aan 19 luisteraars aangeboden in een classificatietaak en een vier-interval-oddity discriminatietaak (4I-oddity). Tijdens de classificatietaak werd telkens één stimulus aangeboden waarna de luisteraar een keuze maakte uit het antwoord /u/ of /i/. In 4I-oddity hoorde de luisteraar vier stimuli waarvan er één verschillend was (de ‘oddball’) en deze oddball was altijd de tweede of de derde stimulus. De luisteraar gaf aan in welk interval de oddball werd waargenomen door op een computerscherm het responseveld aan te klikken met het getal “2” of “3”. De 4I-oddity taak was gekozen in de veronderstelling dat luisteraars zouden kunnen kiezen uit twee discriminatiestrategieën: 1) de luisteraar kan de oddball ontdekken door de akoestische informatie in de stimuli te vergelijken (auditieve codering); 2) de luisteraar kan de oddball ontdekken door te bepalen welke klinkers hij hoort (foneemlabelling). Deze taak werd gekozen omdat van een aantal meer traditionele taken het vermoeden bestond dat de structuur van de taak de luisteraar stuurt naar een exclusief gebruik van foneemlabelling. Aangezien de foneemlabellingstrategie altijd wordt toegepast tijdens de classificatie van spraakstimuli, betekent dit dat de discriminatie- en classificatieresultaten eerder zullen overeenkomen. Je zou kunnen zeggen dat sommige discriminatietaken een bias hebben voor categorische perceptie. Met de keuze van de 4I-oddity taak werd getracht deze bias te vermijden.

Was er een verschil in de mate van categorische perceptie tussen de woord- en tekstklinkers? Nee, de resultaten voor de twee klinkerreeksen waren vrijwel identiek: er was in beide gevallen geen sprake van een relatie tussen discriminatie en classificatie en dus ook geen sprake van categorische perceptie. De discriminatieresultaten waren zelfs lager dan voorspeld werd door de classificatieresultaten. Er werd een controle-experiment uitgevoerd waarin de verschillen tussen de stimuli groter waren om er zeker van te zijn dat het ontbreken van de relatie tussen discriminatie en classificatie niet veroorzaakt werd door een te hoge moeilijkheidsgraad van de discriminatietaak. Dit was niet het geval: de nieuwe discriminatiescores waren hoger, maar er was nog steeds geen relatie tussen de classificatie- en discriminatieresultaten: er was een piek (de foneemgrens) in de classificatiecurve, terwijl de discriminatiecurve bestond uit een vlakke lijn. Dit betekent dat de 4I-oddity taak de luisteraars stimuleerde vooral de auditieve codering te gebruiken en geen foneemlabelling. Dit is niet in overeenstemming met de categorische-perceptie-hypothese volgens welke luisteraars niet in staat zijn iets anders te gebruiken dan foneemlabels bij de perceptie van spraakstimuli. Deze resultaten duiden op een sterk effect van de discriminatietaak op de mate van categorische perceptie. Dit effect werd nader onderzocht in het volgende experiment.

Het effect van de discriminatietaak op categorische perceptieresultaten werd uitgediept in het experiment dat beschreven wordt in hoofdstuk 3. Zoals gezegd is er een vermoeden dat de structuur van de discriminatietaak bepaalt welke perceptieve strategie de luisteraar toepast tijdens discriminatie. Bij de 4I-oddity taak bleken luisteraars exclusief gebruik te maken van de auditieve codering: ze baseerden hun beslissing op de akoestische verschillen tussen de stimuli. In het nieuwe experiment werden de resultaten van 5 discriminatietaken vergeleken: AX, 2IFC, AXB, 4IAX en 4I-oddity. Het doel van het experiment was te zien of de mate van categorische perceptie varieert als een functie van de discriminatietaak. Er werd verwacht dat er een hoge mate van categorische perceptie zou zijn voor de 2IFC-taak (alleen foneemlabelling), en een geringe mate van categorische perceptie voor de vier-intervaltaken (alleen auditieve codering). De mate van categorische perceptie voor AX en AXB zou hier tussenin moeten liggen, omdat luisteraars tijdens deze taken beide strategieën zouden kunnen gebruiken. Een trial in de AX taak bevatte twee stimuli in de volgende combinaties: AB, BA, AA of BB. De luisteraar oordeelde of de stimuli identiek of verschillend waren. In 2IFC hoorde de luisteraars twee verschillende stimuli, AB en BA. De luisteraar gaf aan wat de stimulusvolgorde was; in dit experiment met klinkers was de volgorde /u-i/ of /i-u/. De AXB taak was een drie-intervaltaak waarin de luisteraar besliste of de tweede stimulus identiek was aan de stimulus in het eerste of in het derde interval. Tijdens de 4IAX-discriminatie hoorde de luisteraar vier stimuli in acht mogelijke combinaties, onder meer ABBB en BBAB. De pauze tussen interval 2 en 3 was langer dan tussen interval 1-2 en 3-4, waardoor de luisteraar de stimuli paarsgewijs waarnam. De luisteraar besliste welk paar twee identieke stimuli bevatte.

Uit de resultaten bleek dat de mate van categorische perceptie sterk varieerde als een functie van de discriminatietaak. Zoals verwacht was er een sterke relatie tussen 2IFC-discriminatie en classificatie, en dus een hoge mate van categorische perceptie. Bij AX en AXB-discriminatie gebruikten luisteraars zowel een directe vergelijking van de auditieve sporen van de stimuli als foneemlabelling: de discriminatie was beter dan voorspeld en alleen bij AX was er een piek die overeenkwam met de piek in de classificatiedata. De resultaten van 4IAX-discriminatie impliceerden dat vooral de auditieve sporen vergeleken waren: er was geen sterke overeenkomst tussen de discriminatie- en classificatieresultaten. De 4IAX-functie vertoonde geen piek op de foneemgrens en discriminatie van de stimuli met hetzelfde foneemlabel was beter dan werd voorspeld door de classificatie.

Naast de verschillen tussen de discriminatietaken, bleken er bij een aantal taken verschillen te zijn tussen luisteraars. Bij AX- en 4IAX-disciminatie was er in de resultaten van sommige luisteraars een relatief groot effect van foneemlabelling en dus een hogere mate van categorische perceptie, terwijl andere luisteraars leken te kiezen voor de auditieve strategie waardoor de relatie tussen hun discriminatie- en classificatieresultaten veel minder sterk was. Dit betekent dat we nog een factor kunnen toevoegen aan het lijstje met factoren die effect hebben op de mate van categorische perceptie: de luisteraar.

In het tweede experiment van hoofdstuk 3 werd het effect van het temporele interval tussen twee stimuli in de discriminatietaak op de mate van categorische perceptie bepaald. Bij een relatief lang interstimulus-interval (ISI) werd verwacht dat de luisteraar zou terugvallen op foneemlabelling, omdat het auditieve spoor is vervaagd. In onze eerste serie experimenten was het ISI 200 ms, omdat in eerder onderzoek was aangetoond dat het auditieve spoor na ongeveer 250 ms is verdwenen. In het huidige experiment werden drie taken herhaald met een ISI dat meer dan twee keer zo lang was, 500 ms. Deze taken waren 2IFC, 4IAX en 4I-oddity discriminatie.

De resultaten waren niet conform onze verwachting. De perceptieve strategie van de luisteraars was niet veranderd als gevolg van het langer durende interval. De discriminatieresultaten voor de drie verschillende taken (2IFC, 4IAX en 4I-oddity) waren gestegen, maar het patroon in de discriminatiefuncties was niet gewijzigd. Dit betekent dat de relatie tussen discriminatie en classificatie en dus de mate van categorische perceptie onveranderd was. Alleen in de 4IAX-resultaten was er een kleine verschuiving in de richting van de labellingstrategie. Deze was niet te zien in de gemiddelde data, maar de discriminatiefunctie van een aantal luisteraars vertoonde nu een piek die niet aanwezig was bij het kortere ISI. De binnen-foneemcategorie-discriminatie bleef onveranderd beter dan voorspeld door de classificatie. Het effect van het ISI op de 4IAX data is te verklaren door de structuur van de taak, die een relatief lang interval bevat tussen de twee stimulusparen die in een trial worden aangeboden.

Het effect van het ISI was minder groot dan verwacht, maar de mate van categorische perceptie blijkt sterk te varieren als gevolg van de discriminatietaak. Afhankelijk van de taak werden dezelfde stimuli soms categorisch waargenomen en soms niet. Dit betekent dat de relatie tussen discriminatie en classificatie niet wordt veroorzaakt door de manier waarop we spraak waarnemen, maar door verschillen tussen taken.

In hoofdstuk 4 wordt beschreven hoe de mate van categorische perceptie werd bestudeerd van twee verschillende foneemcontrasten, plofklanken en klinkers. Uit de literatuur blijkt dat de mate van categorische perceptie varieert als een functie van het foneemcontrast dat getest wordt. Plofklanken worden het meest categorisch waargenomen, gevolgd door fricatieven, nasalen, liquidae en klinkers. De resultaten van een recent onderzoek doen echter vermoeden dat dit verschil in de mate van categorische perceptie kleiner wordt wanneer geen synthetische, maar natuurlijke stimuli gebruikt worden. In het huidige experiment werd de mate van categorische perceptie van natuurlijke plofklanken en tekstklinkers (zie hoofdstuk 2) onderzocht. Naast het effect van stimulusnatuurlijkheid werd het effect van de discriminatietaak bepaald. In de voorafgaande experimenten hadden we al laten zien dat de discriminatietaak een systematisch effect heeft op de mate van categorisch perceptie van klinkers. Het effect van de taak op de categorische perceptie van plofklanken is niet eenduidig: soms varieert de discriminatie als een effect van de taak, maar in andere experimenten lijkt het effect van de taak op de discriminatie van plofklanken kleiner dan op die van klinkers. Om het effect van de discriminatietaak te testen op de mate van categorische perceptie van klinkers en plofklanken werden 2IFC en 4IAX uitgevoerd.

De resultaten van de plofklanken en de klinkers kwamen sterk met elkaar overeen. Voor beide foneemcontrasten was er een effect van de discriminatietaak op de mate van categorische perceptie. Voor de klinkers was dit al aangetoond in de voorafgaande experimenten. Voor zowel de klinkers als de plofklanken was de mate van categorische perceptie sterker wanneer discriminatie getest werd met de 2IFC-taak dan met de 4IAX-taak. Het patroon in de 4IAX-resultaten bevatte geen significante pieken en dalen, waardoor de discriminatie van de stimuli met hetzelfde foneemlabel veel beter was dan voorspeld werd door de classificatieresultaten, ook voor de plofklanken. Concluderend blijkt uit deze resultaten dat ook de categorische perceptie van plofklanken geen robuust verschijnsel is en dat met name de discriminatietaak, en daaraan inherent de beslissingsstrategie van de luisteraars, bepaalt of de resultaten categorisch zijn of niet.

De ontwikkeling van de categorisatie van fonemen werd onderzocht met een experiment naar de verschillen in weging van akoestische spraakcues door kinderen en volwassenen. Dit staat beschreven in hoofdstuk 6. Zoals eerder is gezegd, is er een grote variatie in het akoestische signaal als gevolg van verschillen tussen sprekers en de invloed van fonemen onderling. Dit betekent dat er vrijwel geen invariante akoestische cues te duiden zijn voor een bepaald foneem. Onderzoek met volwassen luisteraars heeft aangetoond dat zij deze verschillende cues integreren om te komen tot een beslissing over een foneemcategorie. Tijdens dit integratieproces worden sommige cues zwaarder gewogen dan andere. In recent onderzoek is de ontwikkeling van de weging van akoestische cues bestudeerd. De resultaten laten zien dat kinderen van 3;6 tot 7 jaar akoestische cues voor fricatieven anders wegen dan volwassen luisteraars: de formanttransitie-cue wordt zwaarder gewogen en de fricatiefruis minder zwaar. Naarmate de taalontwikkeling vordert verschuift het wegingspatroon richting het patroon van volwassenen luisteraars.

Het huidige onderzoek had als doel de fricatiefresultaten te repliceren met natuurlijk stimulusmateriaal en het wegingsonderzoek uit te breiden met twee andere foneemcontrasten: plosieven en klinkers. De cues voor de plosieven waren de formanttransitie en de ruisplof en voor de klinkers het spectrum en de duur. De luisteraars waren kinderen van 4, 6 en 9 jaar en volwassenen. Het stimulusmateriaal voor het fricatiefcontrast in de woorden /💧🏳&/ en /♦🏳&/ bestond uit een continuüm tussen de ruis van /💧/ naar /s/ in 6 stappen. De 7 stukken ruis werden vervolgens samengevoegd met het /-🏳&/ gedeelte uit het woord /💧🏳&/ en dit werd herhaald met het /-

⚐&/ deel uit het woord /⧫⚐&/. Zo ontstonden twee stimulusreeksen met twee verschillende formanttransities in de klinker. Op deze manier werden ook de stimulusreeksen gegenereerd voor de plosieven in /◻⚐◻/ en /&⚐◻/. Voor de klinkers in /⌘✌&/ en /⌘♋⑥&/ werd een continuüm gemaakt door interpolatie van het klinkerspectrum en de duur van de klinkers was in het ene stimuluscontinuüm die van de /✌/ en in het andere die van de /♋⑥/. De stimuli werden aangeboden in een classificatietaak waarbij de kinderen en volwassenen na elk woord een keuze maakten uit twee plaatjes.

Uit de resultaten bleek dat de volwassen luisteraars de fricatieven konden herkennen op basis van de ruisinformatie alleen, de formanttransitie-cue werd niet gebruikt. Bij de plosieven en de klinkers was er wel sprake van cue-integratie: volwassenen gebruikten zowel de ruisplof als de formanttransities voor de classificatie van de plofklanken, en zowel het spectrum als de duurinformatie voor de classificatie van de klinkers. Daarnaast toonden de classificatieresultaten aan dat kinderen de akoestische cues anders wegen dan volwassenen. Het grootste verschil was er bij de wegingsschema’s van de vierjarigen kinderen: zij wogen de formanttransities zwaarder en de ruiscue minder zwaar dan de oudere kinderen en de volwassenen. De kinderen van 6 en 9 jaar wogen de ruisplof voor de plofklanken ook minder zwaar dan de volwassenen. In de classificatie van de klinkers waren de leeftijdsverschillen minder groot. Er was een significant verschil in de weging van de spectrale informatie door de 4- en 6-jarige kinderen: ze wogen deze cue minder zwaar. De duur-cue werd door alle luisteraars even zwaar gewogen. De wegingsresultaten laten zien dat de ontwikkeling van het classificatiesysteem nog niet voltooid is op negenjarige leeftijd. Daarnaast bevestigen deze resultaten de bevindingen van de voorafgaande experimenten met volwassenen. Kinderen van 4, 6 en 9 jaar blijken hun weging van cues voor fonemen op subtiele wijze aan te passen. Als de foneemclassificatie gepaard zou gaan met een verslechtering van het discriminatievermogen, dan is het niet aannemelijk dat die weging nog zo veranderlijk kan zijn. De ontwikkeling van de weging van akoestische cues maakt het onwaarschijnlijk dat het auditieve systeem van deze kinderen tegelijkertijd slechter wordt in het onderscheiden van kleine akoestische verschillen in spraak die minder spraakrelevant zijn.

Conclusie

Uit dit onderzoek is gebleken dat categorische perceptie, gedefinieerd als een strikte relatie tussen discriminatie en classificatie van spraakklanken, sterk varieert als een effect van verschillende keuzen in de methode van onderzoek, vooral als een effect van de keuze van de discriminatietaak. Wanneer de discriminatie van spraakklanken wordt getest met een taak die gevoelig is voor de akoestische verschillen tussen de stimuli, is er geen relatie tussen discriminatie en classificatie. Dit betekent dat de ontwikkeling van het perceptieve classificatiesysteem (categorische perceptie in brede zin) geen nadelige invloed heeft op de manier waarop luisteraars spraak waarnemen. Het auditieve systeem blijft intact, ook voor de discriminatie van verschillen tussen verschillende realisaties van hetzelfde foneem of de discriminatie van fonemen die niet in de moedertaal voorkomen. Het onderzoek naar de ontwikkeling van de weging van akoestische spraakcues toont aan dat het classificatiesysteem nog niet volledig verworven is op negenjarige leeftijd. Daarnaast indiceren de vrij subtiele verschillen in weging van de geselecteerde cues nogmaals dat categorische perceptie niets te maken heeft met het afleren van het waarnemen van subtiele verschillen in het spraakgeluid die optreden als een gevolg van verschillende realisaties van een foneem. Integendeel, het categorische perceptie mechanisme zorgt voor een snelle en efficiënte verwerking van spraak, gericht op de herkenning van de klasse van fonemen die specifiek is voor de moedertaal.

Curriculum Vitae

Ellen Gerrits was born in Helenaveen on December 10th 1967. She attended Boschveld College in Venray and obtained her Atheneum diploma in 1986.

In 1990 she graduated as a speech therapist and started working at the neurology department of a regional hospital and in a private practice. From 1991 to 1994 she combined this work with the study of Speech and Language Pathology at the University of Nijmegen. She received her Master's degree in 1995 and in that year she was also employed as a research-assistant at the Max Planck Institute for Psycholinguistics in Nijmegen.

In 1996 she became affiliated to the Utrecht Institute of Linguistics OTS of Utrecht University. As a PhD student (AiO) she conducted the research that is described in this dissertation. During this period she was the PhD-representative for the Dutch Graduate School of Linguistics (LOT).

Currently, she is a post-doc researcher at the Utrecht Institute of Linguistics OTS and, as such, is exploring the speech perception of children at risk of dyslexia and of children with specific language impairment.