Janne Savela

Role of Selected Spectral Attributes in the Perception of Synthetic Vowels

Turku Centre *for* Computer Science

TUCS Dissertations
No 119, June 2009

Role of Selected Spectral Attributes in the Perception of Synthetic Vowels

Janne Savela

To be presented, with the permission of the Faculty of Mathematics and Natural Sciences, for public criticism in the Beta Auditorium at Department of Information Technology in University of Turku on June 26th, 2009 at 12:00 p.m.

University of Turku
Department of Information Technology
FI-20014 Turku, Finland

2009

Supervisors

Professor Tapio Salakoski, PhD
Department of Information Technology
University of Turku
Turku, Finland

Professor Olli Aaltonen, PhD
Department of Speech Sciences
University of Helsinki
Helsinki, Finland

Reviewers

Professor Åke Hellström, PhD
Department of Psychology
University of Stockholm
Stockholm, Sweden

Professor Unto K. Laine, PhD
Laboratory of Acoustics and Audio Signal Processing
Helsinki University of Technology
Espoo, Finland

Opponent

Head of Laboratory Einar Meister, PhD
Laboratory of Phonetics and Speech Technology
Tallinn University of Technology
Tallinn Estonia

ISBN 978-952-12-2308-2
ISSN 1239-1883

Turku 2009

Abstract

This thesis is an experimental study regarding the identification and discrimination of vowels, studied using synthetic stimuli. The acoustic attributes of synthetic stimuli vary, which raises the question of how different spectral attributes are linked to the behaviour of the subjects. The spectral attributes used are formants and spectral moments (centre of gravity, standard deviation, skewness and kurtosis). Two types of experiments are used, related to the identification and discrimination of the stimuli, respectively. The discrimination is studied by using both the attentive procedures that require a response from the subject, and the pre-attentive procedures that require no response. Together, the studies offer information about the identification and discrimination of synthetic vowels in 15 different languages. Furthermore, this thesis discusses the role of various spectral attributes in the speech perception processes. The thesis is divided into three studies. The first is based only on attentive methods, whereas the other two concentrate on the relationship between identification and discrimination experiments. The neurophysiological methods (EEG recordings) reveal the role of attention in processing, and are used in discrimination experiments, while the results reveal differences in perceptual processes based on the language, attention and experimental procedure.

Acknowledgements

The thesis was carried out at several academic departments: Department of Information Technology, Department of Phonetics and Centre of Cognitive Neuroscience in University of Turku and under support of the Finnish school of Language Technology.

I thank my supervisor Professor Tapio Salakoski PhD in Department of Information Technology (University of Turku).

I thank my supervisor Professor Olli Aaltonen PhD, in Department of Speech Sciences (University of Helsinki).

I thank my reviewers, Professor Åke Hellström PhD and Professor Unto K. Laine PhD.

Furthermore I thank my former supervisor, Professor Sirkka Saarinen PhD, in Department of Fenno-Ugric linguistics (University of Turku).

In field of neurosciences, I thank Professor Heikki Hämäläinen in Centre of Cognitive Neuroscience at University of Turku. Furthermore, I thank Professors Risto Näätänen and Heikki Lang in their role in the process of the thesis. I also thank Jyrki Tuomainen PhD, Teija Kujala PhD. Furthermore, I thank Maria Ek MSc, Tuulia Kleimola MA and Leena Mäkelä MA in their role of in experimental research. In phonetics, I thank Ilkka Raimo PhilLis, Markus Mattila MSc Outi Tuomainen MA and other co-workers in the Turku Vowel Test process. I thank also Heidi Lehtola MA, Stina Ojala PhilLis and Lotta Alivuotila MA. In statistics, I thank, Prof Esa Uusipaikka PhD and Jari Lammela MSc.

Finally I thank my parents Arno and Kaarina, my brothers Jussi and Antti and all relatives who have supported me during the years.

Turku 29.5.2009

Contents

List of Figures

List of tables

Articles related to thesis

Aaltonen, O., Hellstöm, A., Peltola, M., Savela, J. and Tamminen, H. (2008). Brain responses reveal hardwired detection of native-language rule violations, Neuroscience Letters 444: 56-59.

Alivuotila, L., Hakokari, J., Savela, J., Happonen, R.-P. and O., A. (2007). Perception and imitation of Finnish open vowels among children, naïve adults and trained phoneticians. 16th International Congress of Phonetic Sciences, Saarbrücken, Universität des Saarlandes.

Alivuotila, L., Savela, J. and Aaltonen, O. (2008). Kielitaustan vaikutus vokaaleja matkittaessa, Puhe ja Kieli/Tal och Språk/Speech and Language 28 (3): 129-140.

Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpaa, M., Kero, P., Raz, A., Kaipio, M. L., Hiltunen, J., Aaltonen, O., Savela, J. and Hamalainen, H. (2002). Speech sounds learned by sleeping newborns, Nature 415 (6872): 599-600.

Eerola, O., Laaksonen, J.-P., Savela, J. and Aaltonen, O. (2003a). Perception and production of the short and long Finnish vowels: Individuals seem to have different perceptual and articulatory templates. 15th Congress of Phonetic Sciences 3.-9. August 2003, Barcelona, Universitat Autònoma de Barcelona: 989-992.

Eerola, O., Laaksonen, J.-P., Savela, J. and Aaltonen, O. (2003b). Suomen [y /i] ja [y: /i:]-vokaalien tuotto havainto kokeiden valossa. Fonetiikan päivät, Akustikan ja äänenkäsittelytekniikan laboratorio, TKK, Otaniemi: 109-113.

Eerola, O., Savela, J., Laaksonen, J.-P. and Aaltonen, O. (2003). Keston vaikutus suomen [y] / [i] jatkumon kategorisointiin ja [i] vokaalien prototyypin havaitsemiseen. Fonetiikan päivät, Akustikan ja äänenkäsittelytekniikan laboratorio,, TKK, Otaniemi:115-122.

Martynova, O., Savela, J., Tuomainen, J., Aaltonen, O., Erkkola, R. and Cheour, M. (2002). Mismatch negativity to natural and synthetic consonant-vowel syllables in neonates and adults. Child Speech Acoustics, St. Petersberg State University.

Martynova, O., Tuomainen, J., Savela, J., Aaltonen, O., Erkkola, R. and Cheour, M. Mismatch negativity to nautral and synthetic consonant-vowel syllables in neonates and adults. submitted

Raimo, I., Savela, J. and Aaltonen, O. (2003). Turku Vowel Test. Fonetiikan päivät, Akustikan ja äänenkäsittelytekniikan laboratorio,, TKK, Otaniemi: 45-52.

Raimo, I., Savela, J. and Aaltonen, O. (2005). Vokaalit testissä. Puheen Salaisuudet. A. Iivonen: 171-181.

Raimo, I., Savela, J., Launonen, A., Kärki, T., Mattila, M., Uusipaikka, E. and Aaltonen, O. (2002). Multilingual Vowel Perception. ISCA Workshop on Temporal Integration in Perception of Speech, 8.- 10. April 2003 France: 86.

Savela, J. (1999). Tutkimuskomisyrjäänin ja suomen vokaalifoneemien rakenteista – A contrastive study on Komi and Finnish vowels, Sanajalka 41: 167-171.

Savela, J. (2000). On the acoustic nature of Komi Zyrian vowels. Congressus Nonus Internationalis Fenno-Ugristarum, Tartu: 158-161.

Savela, J., Ek, M., Kujala, T., Lang, A. H., Aaltonen, O. and Näätänen, R. (2000). Comparison of two vowel systems by mismatch negativity. Neuroscience 2000 Finland Symposium, Helsinki.

Savela, J., Kleimola, T., Mäkelä , L., Tuomainen, J. and Aaltonen, O. (2003). Distinktiivisten piirteiden vaikutus vokaalien havaitsemiseen tietoisella ja esitietoisella tasolla. Fonetiikan päivät, Akustikan ja äänenkäsittelytekniikan laboratorio, TKK, Otaniemi: 39-44.

Savela, J., Kleimola, T., Mäkelä, L., Tuomainen, J. and Aaltonen, O. (2003). The effects of distinctive features on the perception of vowel categories. the15th International Congress of Phonetic Sciences, Barcelona, Spain, Universitat Autònoma de Barcelona: 1000–1003.

Savela, J., Kujala, T., Ek, M., Tuomainen, J., Aaltonen, O. and Näätänen, R. (2001). Suomen ja komin vokaalit esitietoisella ja tietoisella tasolla. 21. Fonetiikan päivät, Turku, Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja: 121-127.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O. and Näätanen, R. (2003). The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages, Cognitive Brain Research 16 (2): 250-256.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O. and Näätänen, R. (2002). Comparison of two vowel systems by the MMN and reaction times. Temporal intergration in the Perception of Speech, Aix-en-Provence, Cambridge University Press, UK:89.

Savela, J., Ojala, S., Aaltonen, O. and Salakoski, T. (2007). Role of different spectral attributes in vowel categorisation: the case of Udmurt. Nodalida 2007, Tarttu, University of Tarttu: 384–388.

Savela, J. and Pikkanen, O. (2005). Role of the higher formants in vowel identification. Studies in Speech perception. M. Peltola and J. Tuomainen. Turku, Department of Finnish and General linguistics: 15-26.

Savela, J., Raimo, I., Uusipaikka, E., Aaltonen, O. and Salakoski, T. (2008). The categorisation of synthetic vowels by Swedish speaking listeners in Finland. Fonetiikan päivät 2008, Tampere: 109-121.

Tuomainen, J., Savela, J., Obleser, J. and Aaltonen, O. Attention modulates the use of spectral attributes in vowel discrimination: Behavioral and event-potential related evidence. MMN 2009 – Fifth Conference on Mismatch negativity (MMN) and its Clinical Applications, April 4-7, 2009, in Budapest Hungary.

1. INTRODUCTION

This study addresses the issue of how people categorise synthetic vowels. Theoretically, there are two fundamental questions regarding the nature of vowel perception, the first of which concerns the structure of vowel categories, or the way in which synthetic stimuli are categorised and what structures the categorical spaces have. The second question pertains to the processes behind the categorisation, that is, the processing aspects of the vowel perception or the temporal processes behind the outcome of the categorisation process. This is, in other words, the process order in the categorisation of a particular stimulus, the way in which the stimulus activates the brain's memory mechanisms. Together, the various processes provide evidence regarding the semiotic nature of vowel perception, the way in which a listener interprets the sounds they perceive. Thirdly, this thesis is about the processing of acoustic features and their role in perceptual processes, that is, the inventory or emphasis of particular acoustic features in a given language.

This study has three main perspectives. Firstly, it compares the identification and goodness ratings of acoustic stimuli in 15 languages. The statistical models of spectral representations (logistic regression models) are used to compare the relationship between stimuli's acoustic manipulations (the changing values of parameters in the presented physical stimuli) and the identification and goodness rating responses (a response to the task in the experimental paradigm). Secondly, the discrimination response of subjects with different language backgrounds (their native language) is studied, and finally, the role of attention (automatic versus non-automatic stages of processing, see Chapter 2) is studied in the processing of acoustic features (formant frequencies or spectral moments). The results provide evidence that questions the way in which the auditory system processes acoustic features (automatic versus non-automatic stages of processing) and which of these are essential in listeners 'linguistic decisions (as indicated by identification or goodness rating paradigms).

This thesis was required to list the possible aspects of language or speech technology to which it could make a contribution and the following paragraphs will attempt to respond to that requirement. In terms of applications (the technological aspect) the thesis provides information to some fields of speech technology, the first of which concerns the models on vowel identifications in some languages (the 15 languages examined in Study I). The thesis provides results on the prototypes and category boundaries in 15 languages. It describes the differences between languages (the loci of prototype in vowel space and boundaries between the vowel categories) and can be used to set language-dependent parameters of synthesis in terms of aspects of perceptual space (on the basis of the same descriptive analysis).

The second application of the thesis (the behavioural aspect) considers the understanding of language learning. The data can be used in order to model the identification processes in different languages (the same 15 languages in

Study I. This type of information is useful for understanding the feedback mechanisms of speech perception, for example the learning of non-native languages (concrete differences between languages). The knowledge on the perceptual strategies of sound perception in different languages can offer information about how the categories should be studied for non-native listeners.

The third application (the general phonetic aspect) of the thesis concerns the general knowledge of the processing of acoustic features reflected in the vowel spectrum. Speech perception is generally accepted to be a result of evolutionary environment (e.g. Lindblom, 2000) and the speech perception models therefore also include an evolutionary or ecological aspect. This thesis also contributes an understanding of the automatic processes affecting signal-noise relationship in the perception of speech sounds in living animals.

This thesis examines attentive and pre-attentive vowel perception. It consists of three studies (henceforth *Study*) that are an independent series of experiments or analyses (henceforth *Experiment* and *Analysis*). The thesis examines the use of two types of acoustic attributes of vowels, *formants and spectral moments*, in the identification and discrimination processes of stimuli. Statistical tests are carried out to demonstrate how those acoustic attributes are used in vowel perception. The results explain, firstly, the attributes in vowel sounds that are used as identification criteria by subjects with different native languages, and secondly, the how spectral moments and formant peak differences affect pre-attentive neurophysiological processes. Thirdly, the results show how the differences between vowel systems can be understood if isolated synthetic steady state vowels are used.

1.1. Perspectives of the thesis

This thesis is an experimental study of the role of acoustic attributes in vowel perception. The three following types of experiments have been used: identification experiments, goodness rating experiments and discrimination experiments. Identification and goodness rating experiments are known as attentive (because the listeners have to concentrate on the task). Discrimination, however, has been studied with both attentive and pre-attentive methods (that is, tests that required an attentive response and those that did not). In pre-attentive experiments, unlike attentive ones, responses are independent of the subject's consciousness, e.g. Näätänen and Winkler (1999).

The relationship between identification responses and acoustic attributes is demonstrated using logistic regression models, a methodology known as "pattern recognition models" that is related to visual pattern recognition models. The listener divides the perceptual vowel space to the regions that have different shapes. This is referred to as symbolic in this thesis because it requires knowledge about the vowel system from a "global point of view", without earlier knowledge about particular sounds. By using vowel charts and

vowel regression models, some general aspects can be found in vowel identification.

In this thesis, vowel identification is investigated with forced choice tasks, in which every vowel must be categorised, even if it does not clearly belong to any category. In goodness rating tests the "goodness" of a vowel, as a representative of the vowel category, was rated on a scale from 1 (very poor) to 7 (very good). The loci of the prototypes and the vowel category boundaries, observed differently, may give evidence regarding the perceptual vowel space in different vowels.

The discrimination of vowels is studied in the thesis with an odd-ball paradigm, in which the deviating stimulus occurs randomly within a sequence of deviant stimuli. That paradigm is generally used in MMN experiments and also in the attentive experiments on similar sound blocks. The MMN is a good tool in terms of temporal resolution, which means that it gives a good picture about the temporal order of different processes.

In odd-ball experiments the standard stimulus is proposed to activate the mental representation of the sound, while a new sound in the stimulus stream creates a mismatch between the standard and the deviant. This mismatch is suggested to reveal the perceptual processing of two perceived sounds. In the attentive task, subjects were asked to press a button when they heard a difference in the vowel stream. In the pre-attentive task, however, subjects listened passively to the same stimuli (they were asked to ignore them), and the neural MMN-responses during the task were recorded. The MMN is arguably an automatic stimulus-specific process in which the memory traces for certain acoustic features create the response to the properties of the new sound.

Study I investigates whether the same acoustic attributes are used in the identification of all vowels within the same category. The results of identification and goodness rating tests are expected to be related to different acoustic criteria because they stem from different perceptual processes. Firstly, the identification responses of subjects with different vowel systems are discussed, and, secondly the possible relationships between spectral moments and the responses are examined by observing the vowel charts for different listener groups.

Logistic regression methods, which demonstrate how the responses are related to the measured acoustic variables of the stimulus set, are used for statistical analysis. Such methods indicate the presence of a statistical dependence between response and the acoustic attributes of stimuli (that is, formant frequencies and/or spectral moment values). The *relative strength* of explaining attributes can be used to describe the categorisation criteria of vowels in particular languages, in terms of linear combinations. The relative strength indicates the importance of each attribute as a categorisation criterion in a particular statistical model.

Analysis 1 of Study I will show that the identification and goodness rating of the Udmurt /u/ -category is based on different acoustic attributes. It will be argued that a simpler model can be used to describe the goodness ratings of vowel stimuli than the categorisation of the same stimuli. *Analyses 2* and *3* of Study I examine the relative strength of different acoustic attributes in two pairs of languages with a fixed number of vowel categories (5 or 8). Formant-based models and whole-spectrum models are compared on the basis of the identification response (*Analysis 2*) and goodness criteria (*Analysis 3*). The study is described in general in Raimo, Savela, Launonen, Kärki, Mattila, Uusipaikka & Aaltonen, 2002; Raimo, Savela & Aaltonen, 2003; Savela, Pikkanen, Raimo, Uusipaikka & O., 2004; Raimo, Savela & Aaltonen, 2005. Analysis 1 of Study I is published in Savela et al. (Savela, Ojala, Aaltonen & Salakoski, 2007).

Study II examines the role of stimulus (as standard or deviant within an odd-ball stimulus block) in the attentive (*Experiment 1*) and pre-attentive processing stages (*Experiment 2*) by comparing the mismatch negativity (MMN) amplitudes and reaction times (RTs). Similar vowels representing typical Komi and Finnish vowels are compared. The aim of Study II is to reveal whether attention affects the discrimination between prototypical and non-prototypical sounds. Furthermore, the role of spectral moments is studied in the context of attentive versus pre-attentive tasks. Study II is published in several versions: Savela, Ek, Kujala, Lang, Aaltonen & Näätänen, 2000; Savela, Kujala, Ek, Tuomainen, Aaltonen & Näätänen, 2001; Savela, Kujala, Tuomainen, Ek, Altonen & Näätänen, 2002; Savela, Kleimola, Mäkelä, Tuomainen & Aaltonen, 2003b; Savela, Kujala, Tuomainen, Ek, Aaltonen & Näätänen, 2003.

Study III compares the discrimination of two continua by using the odd-ball paradigm in both attentive (*Experiment 2* and *3*) and pre-attentive (*Experiment 4*) experiments. In *Experiment 1,* stimuli are selected for *Experiments 2, 3* and *4,* while in *Experiments 2* and *4* the vowel difference is fixed in formant but not fixed in CoG space. The difference between languages in the use of spectral attributes is tested in *Experiment 2* with Spanish subjects as a control group in order to test the role of language in the discrimination of stimuli. In *Experiment 3* the CoGs are manipulated by inserting white noise into the vowels, which affects the distance between CoGs. In Experiment 4 the MMN are recorded for same stimuli as in Experiment 2.

The hypothesis of Study III is that the automatic discrimination utilises formants, whereas the attentive discrimination also utilises whole spectrum attributes. The formants provide more indexical information about vowels (see Chapter 2.1), whereas the use of the spectral moments is dependent on attention and on the ability to focus on the iconic (for definition, see Chapter 2.1) features of the stimulus. The hypothesis is expected to be true if CoGs, as compared to the formants, contribute to a difference in RTs (attentive processing) but not in MMNs (pre-attentive processing). If so, this would

indicate that formants are the features used in the buffering stage and in forming the pre-attentive auditory representations of vowels. Such dissociations have not been reported previously in the literature. The results of Study III are published in Savela, Kleimola et al., 2003b; Savela, Kleimola, Mäkelä , Tuomainen & Aaltonen, 2003a.

Together, the three studies show which spectral attributes are reflected in the identification and discrimination performance.

2. VOWEL PERCEPTION AS A THEORETICAL ISSUE

2.1. Semiotic foundations of language

The semiotic foundations of language refer to the relationship between the perceived sounds and the cognitive categories. The question is how, in the mind of the listener, a particular sound is linked to the given category. For a listener, sound induces many reactions, which tell the experimenter about the nature of the cognitive processes related to speech. This paragraph is interested in the semiotic nature of speech perception, which means the nature of vowel categories as interpretants, or their relationship to the other concepts in mind of the perceiver. The emphasis of semiotics is important because it used to be important when formulating the theories on categorical structure of speech perception (e.g. Jakobson, Waugh & Taylor, 1979). Many of the key concepts in phonology stem from this debate (phonetic versus phonological representation). This thesis employs one semiotic tradition that disintegrates the levels of abstract concepts and the familiarity-based internal representations as different types of semiotic communication. The aim of this chapter is to review the literature of speech perception studies by concentrating on the following question: Is speech perception based on the holistic system of symbolic categories, like in phonemic classification of the speech signal (holistic semiotic view), or on learning to control the speech mechanisms and to conceptualise the speech signal on the basis of symbolic categories (bottom-up approach).

In one theory concerning the semiotic aspects of human communication, the recognition of sounds can be described as occupying three different levels, *iconic, indexical and symbolic*, which describe any kind of communication of individual listeners in their social environment (Deacon, 2003). The levels that have interested speech researchers until recently are the iconic and the symbolic levels. The iconic level is considered to be the concrete, acoustic/auditory level of sound perception, neutral to experience (e.g. Harnad, 1987), while the symbolic level of speech perception is considered to be abstract and conceptual. At the symbolic level, the distinctive system of symbols defines the status of speech sounds in the sound system (a classical holistic view of the semiotics of language) (Saussure, 1919) (Jakobson, Waugh et al., 1979) (Lotman, 2002). The sound categorisation at the symbolic level is intended to describe the language as a system of symbols described in terms of their relationships with other symbols. It is problematic, therefore, if the continuously changing system of living language is studied (for discussion, e.g. Ravila, 1967).

The decoding approach traditionally considers speech to be an efficient code. According to this view, the question of speech perception can be seen as a question of perceiving a static code (e.g. Tobin, 1990). The idea of static code can be exploited in the studies of information processing of speech (e.g. Pisoni & Luce, 1987). The oldest speech perception theories in cognitive psychology

link acoustic and symbolic capacities to the perception of the "speech code" in speech signals (e.g. Liberman, S., Shankweiler & Studdert-Kennedy, 1967). The first explicit theories of speech sound perception were only presented relatively recently, due to the fact that theoretical and experimental knowledge of speech perception only developed after the Second World War. Speech mode approaches, like the motor theory of speech perception (Liberman, S. et al., 1967), considered the problem of continuous (iconic) categorisation of speech concepts to be an irrelevant issue. The idea was that because contemporary linguistic theories were concentrated on the minimal code describing linguistic competence, speech perception reflected similar processes. According to this approach, the speech particles name themselves; every acoustic stimulus is processed into a canonical linguistic concept (Studdert-Kennedy, 1974). The question that arises from this perspective regards the choice of the perfect semiotic model on the interpretative structures behind the sounds. It was evident that the phonetic percept does not reflect the orthographic versions of the similar linguistic targets, which led to the creation of the theory of hidden linguistic structure behind the words. The continuous phonetic signal was mapped to the abstract entities that could be manipulated in the context of the inner model of the language, which, in the case of phonetics, were ariticulatory/acoustic units. So-called "name-based models" offered a contrasting view of the issue, (e.g. Cross, Lane & Sheppard, 1965). The knowledge of the labels determines the identification function in a particular experiment in these models.

The issue has proven more complicated than first expected. Experimental investigation of the simple rigid link between neutral acoustic/auditory percepts (icon) and abstract conventional symbols (a basic unit of the holistic/phonetic code) that can be used to describe the linguistic competence of the perceiver has been difficult. This difficulty is a result of the fact that there is no transparent correspondence between acoustic similarity, articulatory events and the conceptual architecture of different sounds (Diehl, Lotto & Holt, 2004). The acoustic features used in speech perception vary between speakers and languages. Neither motor gestures nor the acoustic similarity of speech sounds explain the differences between sounds systems without considering the aspects of personal experience and habits of the speaker/listener (e.g. Bell-Berti, Raphael, Pisoni & Sawusch, 1979).

Speech perception has traditionally been studied with attentive experiments. The problem with these experiments is that different modes (iconic, indexical, symbolic) of vowel processing have mixed responses. It is difficult to examine the role of different spectral attributes without considering the effects of post-perceptual (that is, decisional) strategies in the processing of sound percepts (e.g. Niemi & Aaltonen, 1986, Rosner & Pickering, 1994). Recent approaches in phonetic research have looked at the pre-attentive level of sound processing, which enables sound patterns to be studied directly without reference to the linguistic concepts of the subject (for review see, e.g., Aaltonen, Eerola, Hellstrom, Uusipaikka & Lang, 1997).

This thesis argues that, instead of describing speech perception as a decoding process based on symbolic particles, phonetic research should focus on the different levels of semiotic recognition principles (iconic, indexical and symbolic) when examining speech perception. These recognition principles must be distinguished from one another in order to describe the linguistic behaviour of the particular subject. The relationship between the signal (physical stimulus) and the concept (the symbolic category that can be replaced with similar units) must be understood in terms of the hierarchy of these semiotic levels (Deacon, 2003), as provided by Peirce (Peirce, Kloesel & Houser, 1992). The semiotic levels are: 1) sound as an auditory presentation of a physical signal, 2) sound as an index of phonetic intention, and 3) sound as a symbolic category (e.g. /a/), perceived according to the phonological code of language.

Deacon follows Peirce's analysis of the logical relationship between a sign (representamen), its effect on the mind of the perceiver (interpretant) and the anticipated entity that the interpretant is expected to represent for the perceiver (object). The relationship between representamen and interpretant can, in turn, be further divided into phenomenal (iconic), dynamic (indexical) and conceptual (symbolic) versions that refer to immediate, experience-based and analytical knowledge of the object (Deacon, 2003, Vehkavaara, 2003). Consequently, in the identification responses, all the levels of identification can be observed simultaneously if they are represented in the mind of the perceiver.

The hypothesis of this thesis is that the structure of categories, as reflected in human linguistic behaviour, follows this hierarchy. The iconic level is the fundamental level of vowel categorisation. Identification is based on the similarity between the sign and its object, and can be grounded on some kind of non-memory-based mechanism between sound and the entities in the perceptual system.

The iconic level also reflects the absolute base-line in perception (that is, what is physiologically perceivable) depending on the resolution of the perceptual system (e.g. the physiology of the ear). In fact, the perceiver's auditory sensitivity may change according to the experimental design. The same acoustic difference may be easily detected in one design and be perceivable in another design (e.g. ABX vs. odd-ball paradigm), (Harnad, 1987). In terms of speech sounds, the features are based on the afferent properties of the tonotopical organisation of the auditory cortex, the anatomical structure responsible for creating auditory representations.

The next level of identification competence is based on subject's indexical experience. Due to the variation in the acoustic properties of the produced items, which are expected to represent the same vowel category (Peterson & Barney, 1952), vowel studies generally propose that vowel categorisation be based on some type of abstraction of the stimuli (singular features (Jakobson,

Fant & Halle, 1961) or phonetic pattern (Remez, Rubin, Berns, Pardo & Lang, 1994)).

The indexical level of categorisation can be based on icon-to-icon abstraction, shared context or the expected relationship between two sounds. According to Deacon (2003), the stability of the perceptual system is based on indexes, while the flexibility of the symbolic system is based on their solidity. There are three characteristics on the indexical level (Deacon, 2003). Firstly, the number of indexical categories is related to the experience of the listener. In terms of speech sounds, the source of experience is either learned speech act control or stored representations of linguistic patterns. Secondly, the indexical relationship between an object and a sign can be based on uncontrolled indecomposable and learned associations (probably as a result of primitive learning processes). Thirdly, the indexical level treats each perceptual category individually although the categorisation process can be guided spontaneously by the relationships between indexes. The locality of indexical categories can lead to a situation where the same acoustic stimuli can be categorised differently depending on the information available to the listener (Assmann, Nearey & Hogan, 1982).

The third level of categorisation is the symbolic one. In contrast to the indexical level, it defines categorisation in terms of an abstract global system, which can be described as a system of logical oppositions in a multi-dimensional space. The symbolic level is based on the ability to use and reproduce a set of indexical representations in terms of grammar and a set of rules (e.g. Aaltonen, Hellström, Peltola, Savela & Tamminen, 2008). According to Deacon, the symbolic level is based on the ability to recognise the global system between individual indexes (ibid.). On the symbolic level, identification is based on the global relationships between symbols specific to each code. Non-prototypical sounds are categorised on the grounds of the multi-dimensional space of indexical properties between symbols (e.g. formant space). Identification is based on the holistic system of indexical features defined by the subject ś linguistic conventions rather than on the relationship between sounds and their stored representations.

The symbolic level of sound perception can be studied by means of explicit global experiments, such as forced-choice identification tasks. Pattern recognition experiments, therefore, are one method for directly studying subjects ’symbolic conventions.

2.2. Fundamentals of vowel processing

This chapter examines the fundamentals of speech perception and will review the literature on speech perception studies.

2.2.1. Vowel spectrum as a physical object – formants and spectral moments

The description of speech in terms of sounds, gestures and symbols dates back to ancient cultures, and is also naturally related to the emergence of phonetic orthography. The terminology of phonetic description was standardised and conceptualised by the IPA (International Phonetic Association) in 1888. In general, phonetics has been described in terms of indexicality, thus articulatory gestures have been the basis of phonetic descriptions (Bell, 1867; Sievers, 1881; Jespersen, 1904).

Vowels have traditionally been described in terms of acoustic/phonetic variables, (e.g. IPA, 1949). The descriptions still have many pseudo-articulatory names for categories (e.g. roundness to describe the shape of the lips; front-back and close-open to describe the tongue position in the mouth) (IPA, 1949). The earliest description of vowels as a vowel triangle originates from Hellwag (1781), who presented a vowel triangle based on reference points of nine different vowels. The vowel triangle demonstrated vowel systems in relation to one another, making the differences between the systems systematically describable.

The most essential theories on acoustic, articulatory and perceptual phonetics were presented only after the Second World War (Joos, 1948; Fant, 1960), as a result of progress in signal processing research. The most important step was the introduction as a scientific method of the spectrogram, which made speech signals visible (Cooper, Liberman & Borst, 1951) and led to a coherent understanding of the physical mode of formants.

In Fant ś acoustic theory of speech (1960), the filter function of the vocal tract can be described as a series of Helmholtz resonators with different resonating capacities. According to this view, altering volumes of bottle-shaped parts and constrictions between them result in different formant configurations. In more detailed analysis, the effects of different parts of the vowel tract, such as the impedance of the lip (based on its size and mass), affect the amplitude of the vowel. The effects of vocal tracts on the amplitudes of the formants have been studied more thoroughly by Laine (1989).

As a physical object, the vowel can also be seen as a distribution of spectral energy to different frequencies of the spectrum. With this approach, the vowel is understood as an amplitude pattern in which the slope of the stimulus affects the identification functions of the vowel sound (Rosner & Pickering, 1994). In terms of spectral shape (the slope of the amplitude) the vowel sound is similar to the different types of sounds and general colour of the sound, whereas the formants can be considered as the control signal that provides information to the speaker/listener.

Formants are considered, in this thesis, as the indexes of vowel identity. They are known to describe the phonetic distance between sounds, as opposed to

the other features of the vowel spectrum, which are related to the general acoustic distance between sounds. The indexicality of formants (or their perceptual equivalents) is based on the subject's implicit knowledge of the vowel. This indicates the relationship between the vowel and the intention, which is largely based on the biological knowledge of the relationship between acoustic signals and phonetic events (Liberman, S. et al., 1967; Rosner & Pickering, 1994; Fowler, 1996; Nearey, 1997). The formant frequencies of vowels vary between languages, a different that is also reflected in the categorisation habits of persons with different linguistic backgrounds. Vowel categorisation in different languages was the focus of the collection of the Turku Vowel Test database (see Chapter 4) (Raimo, Savela et al., 2003).

Although formants are still the most common acoustic measure of the phonetic similarity between sounds, due to their indexical nature, an alternative measure has also been proposed, by which the vowel may be identified on the grounds of its spectral shape (Bladon & Lindblom, 1981). This approach is known as the *whole spectrum approach* and it has been supported by the following arguments:

1) The traditional formant-based model concentrates only on the two prominent formant peaks, F1 and F2, ignoring the other acoustical aspects of the speech signal (such as differences in amplitude). Instead, in the whole spectrum model, all frequencies affect the quality of the vowel.

2) Formant peaks are not always traceable in the speech signal. An important question is whether they even exist in every signal.

3) Dissociation between perceptual and formant-based similarities between two sounds has been found. Equidistant formant values are not always reflected as the same perceived similarity between two vowel sounds.

Different whole spectrum measures (that is, spectral moments) have been examined in the literature. The spectral moments in focus in this thesis are: centre of gravity, standard deviation, normalised skewness and normalised kurtosis. The spectral moments were chosen to describe the acoustic quality of vowels because of the non-continuity of formant frequency values in some areas of the formant space (see, for example, the relationship of identification results and F2 values in Figure 4) This result provides evidence against the traditional formant approach, which suggest that vowels can be identified simply on the grounds of the first two formants (F1 and F2). In fact, this phenomenon had already been discovered in the 1970s in the vowel colour estimation experiments conducted by Butcher (1974). In Butcher's experiments, English subjects were asked to evaluate cardinal vowels in terms of their colour (*darkness* or *brightness*). In terms of vowel colour, the difference between the back rounded vowel [u] and the front rounded vowel [y] was reported to be smaller than the difference between [u] and [ɨ], although the distance relation in the formant space was opposite in the same

pairs. In other words, the distance in Hz was smaller between [u] and [ɨ] than between [u] and [y]).

The spectral moments, statistical measures based on the distribution of amplitudes over the frequencies of vowel spectrum, are based on both the speaker-dependent features of the glottal source and the filter functions of the vocal tract. Because the amplitude pattern of the final signal reflects both aspects of the speech signal, the spectral moments combine the spectral amplitude pattern of the glottal source and the filter resonances of the vocal tract (Milenkovic & Forrest, 1988).

The spectral moment can also be applied to consonant sounds. For example, Forrest et al. (1988) have recognised the role of spectral moments as an important acoustic determinant in the quantitative analysis of fricatives. The use of spectral moments is introduced in the vowel identification and discrimination experiments conducted in this thesis. The TVT database is used in the identification experiments and the discrimination experiments are run with a separate set of synthetic vowels. A detailed description of spectral moments using the PRAAT analysis programme is described in Study I.

2.2.2. Models in processing speech spectrum

The model of vowel processing followed in this thesis is explained in Figure 1. In some aspects the model resembles the auditory model presented by Rosner and Pickering (1994). In their model the spectrum is transformed into the auditory loudness density pattern (ALP), which in turn is transformed into the phonetic loudness density pattern (PLP), a pattern based on special mechanisms that compute the peaks of the ALP. The transformation results in two peaks of effective vowel indicators with shoulders in between. These two peaks provide the basis for the local vowel indicators (i.e. effective formants) that contribute to the final representation of the vowel. The final representation is then compared with the stored exemplars of a particular vowel category by using a template matching procedure. Different speaker groups, such as males, females and children, have different prototypes. Rosner says that the vowel space for a speaker is always linear and follows the nearest prototype rule. According to this rule, the nearest prototype of a particular vowel category is identified in terms of E1 and E2 space (the perceptual equivalents of F1 and F2). The formant values presented in this thesis are actually E1 and E2 because they are believed to be effective in terms of vowel identification.

In addition to models based on the computational model of speech perception, this thesis is also concerned with psycho-physiological evidence of sound processing. The adequacy of the whole spectrum model (the iconic auditory pattern) and the formant model (the indexical auditory pattern) can also be evaluated in the neural indexes of sound processing. The spectrum of the

sound is processed in the brain and these processes can be examined with EEG measurements (Näätänen & Winkler, 1999). A variety of experimental set-ups can be used in EEG studies to reveal the properties of a sound system, for example, the mismatch negativity component (MMN) in event-related potentials (ERP). The MMN examines the stage at which the pre-attentional activity forms an auditory representation. The stages at which an auditory representation becomes a recognisable speech sound can be examined by other components of the EEG response (e.g. P3b) (Sussman, Kujala, Halmetoja, Lyytinen, Alku & Näätänen, 2004).

Näätänen and Winkler have proposed a model of neural processing of various kinds of sounds (1993; 1999). According to the model, the stimulus is first transformed into an afferent activation pattern (about 100 ms after the stimulus onset). It is then buffered into pre-representational feature traces, and finally transformed into a fully auditory stimulus representation that can be controlled voluntarily. The buffering stage is pre-conceptual, stimulus-specific and relatively long in duration (60 ms). In speech sounds, this can be considered to be the basis of phonetic perception of vowel stimuli. The memory traces stored in the auditory cortex include all the crucial information about sounds (Winkler, Reinikainen & Näätänen, 1993).

The first stage in processing speech sounds is the activation of the corresponding stimulus pattern in the auditory cortex. Some features can be extracted from the stimulus at this stage on the basis of memory traces (e.g. formants). They are reflected in the amplitude changes of ERP when using vowel stimuli (Aaltonen, Niemi, Nyrke & Tuhkanen, 1987; Kraus, 1992; Aulanko, Hari, Lounasmaa, Näätänen & Sams, 1993; Aaltonen, Eerola, Lang, Uusipaikka & Tuomainen, 1994; Shestakova, Brattico, Huotilainen, Galunov, Soloviev, Sams, Ilmoniemi & Näätänen, 2002; Nenonen, Shestakova, Houtilainen & Näätänen, 2003). In vowel identification experiments, the formant peak frequencies have contributed a larger difference in MMN amplitudes than the changes in the whole spectrum (Jacobsen, Schroger, Horenkamp & Winkler, 2003).

The use of different attributes (i.e. measures) in this thesis is reflected in the processing of vowel timbres at different stages (Figure 1). From a perceptual point of view, the attributes can be roughly divided into phonetic (formant-based) and general auditory (both formant- and spectral shape-based) attributes, both of which can code the phonetic quality of sounds. They can be described by comparing the identification and discrimination functions of sounds (Carlson & Granström, 1979; Bladon & Lindblom, 1981).

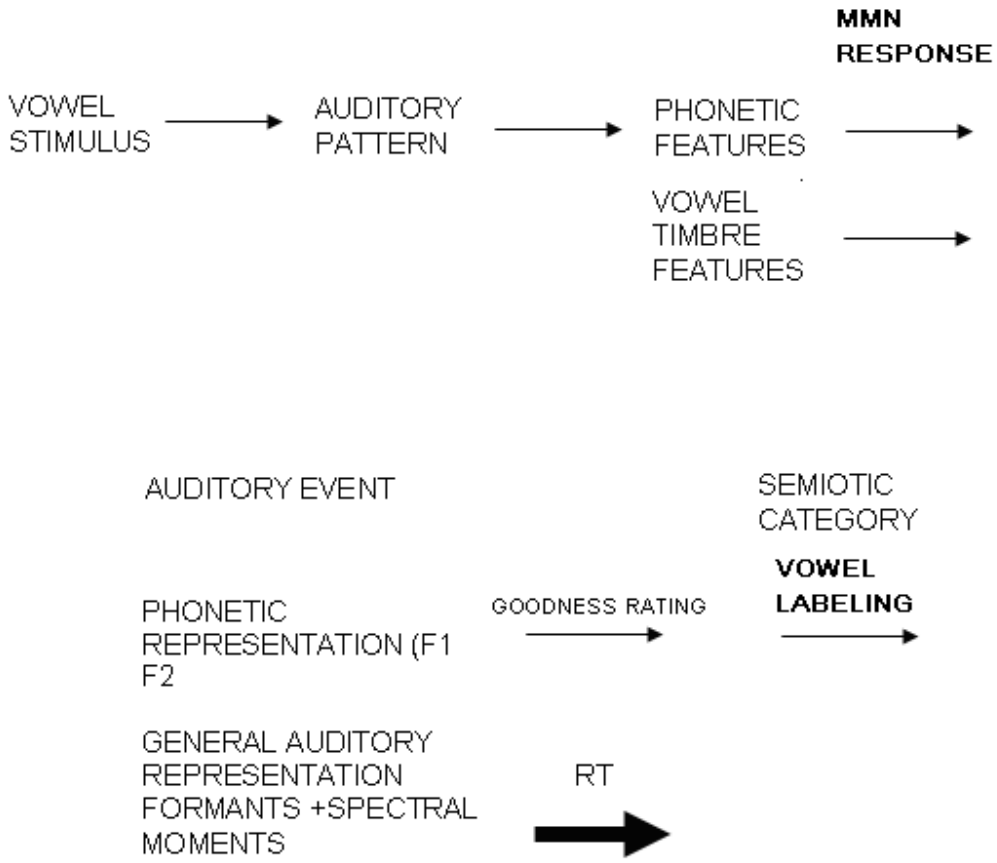


Figure 1. The role of spectral features at different stages of the categorisation process
Bold arrows indicate attentive responses and narrow ones indicate pre-attentive phonetic responses. Terms above the arrows describe a method for demonstrating the particular process.

This thesis claims that the difference between spectral attributes is reflected by the level of conscious sound processing. Therefore, spectral attributes are used to study the identification criteria of synthetic vowels in the same way that acoustic attributes are used in discrimination tasks. According to Schouten, any continuous acoustic attribute within a stimulus set can be used as a discrimination criterion (Schouten, Gerrits & van Hessen, 2003). This thesis expands on Schouten's idea to apply it to identification experiments as well. In forced choice recognition tasks, any acoustic attribute can be used as an identification criterion. The role of different spectral attributes in auditory representations can be demonstrated by comparing the attributes among the vowels in which the systematic variation occurs.

Four types of test designs have been used in this thesis to study the utility of acoustic attributes (formants versus spectral moments) in vowel processing: pre-attentive discrimination tasks, attentive discrimination tasks, identification tasks and goodness rating tasks. These experiments enable the emergence of the subject's vowel categories to be examined from neural responses of acoustic stimuli to more cognitive and social aspects of vowel processing.

2.3. Review of studies - perceptual vowel space

This chapter reviews the current body of knowledge on vowel stimuli and the differences in studies revealing the use of different types of spectral similarity. Firstly, the various acoustic representations are compared in order to dissociate the vowel spaces based on formants and on the whole spectral space (corresponding to iconic versus indexical similarity). Secondly, the possible categorisation criteria are examined by reviewing identification experiments (symbolic versus indexical categorisation).

2.3.1. Auditory patterns of the vowel stimuli

Many experimentally-based scales have been proposed to describe the effects of auditory system, including the *Mel scale* (e.g. Stevens & Volkmann, 1940; Fant, 1973), the *Bark scale* (Zwicker, 1961), and the *ERB scale* (Moore, Peters & Glasberg, 1990). According to each of these scales, the frequency difference between two points in a spectrum is perceived in a linear fashion until 1 kHz and logarithmically thereafter. The Mel scale is based on the perceived similarity of two tones, whereas the Bark and ERB scales are based on the way that critical bandwidths affect the perception of a sound in the ear. The ERB is based on the Roex filter shape derived by Patterson (Patterson, Nimmo-Smith, Weber & Milroy, 1982) from masking data. Roex filter skirts descend gradually from its centre frequency. The ERB measure, which is based on the cumulative number of equivalent rectangular filters as a function of frequency, has been adopted by many studies (e.g. Rosner & Pickering, 1994) and has a better experimental basis than the Bark scale.

Studies on the pre-attentive processing of vowel stimuli have provided a detailed picture of the processing of acoustic properties of the stimuli (for review, see Eggermont, 2001), and have revealed the architecture of neural populations on the auditory cortex coding specific spectral features. The architecture is based, in general, on iso-frequency bands in the auditory cortex that reflect the frequency, bandwidth and asymmetry of the prominent spectral peaks in the signal (Versnel & Shamma, 1998; Diesch & Luce, 2000). Having passed the auditory cortex, the signal is transformed into a combination of prominent features of the spectrum. This combination of features acts as an input for higher mental processes. Ohl et al. showed that cortical representations of vowels indicated by ERP reflect combinations of the lowest

formant values rather than independent formant values (1997). Critical comparisons between formant space and whole spectrum space were not provided.

2.3.2. Iconic space in vowel processing – reflections of alternative spectral attributes

Several iconic methods have been developed in order to study the role of alternative spectral attributes in vowel perception, in which no absolute categories are required Differences between spectral attributes can be examined by using just noticeable difference (JND) measurements. Odd-ball paradigms are an alternative method for studying the formant asymmetries that result from differences in spectral shape. These methods make it possible to obtain language-independent knowledge.

Some interesting results have been found through JND. The lack of perceptual equivalence in formant space has been the expected cause of discrimination differences, so the perceptual difference between vowels mirrors the Euclidean formant distance between them. Perceptual equivalence results in symmetrical JNDs (Flanagan, 1955; Nord & Svantelius, 1979; Hawks, 1994).

In the study of Hawks (ibid), asymmetries were thoroughly examined using sets of vowel stimuli in which the two lowest formants moved in parallel and opposite directions around prototypical cardinal vowels and in the area between them (the area of non-prototypical vowels). The results indicated, in general, that the ability to detect differences in vowel spectra differed between vowels. The prototypicality of sound did not affect the difference limens of a particular sound, however, the direction of formant changes (parallel or opposite) did affect the JNDs. Although parallel movements of F1 and F2 were perceived more accurately than opposite movements, the difference varied in size in the various areas of the vowel space. These results raise the possibility that the spectral shape is affected in vowel discrimination, given that the spectral shape may differ in terms of symmetry from the simple formant-based model. This aspect is not controlled in these studies.

2.3.3. Indexical level of vowel perception – the role of experience within vowel perception

In recent years, new theories of vowel categorisation have emerged, in parallel with general psychological theories that consider all kinds of categorisation. Knowledge about categorisation has developed from complicated analytical theories (containing a lot of advance information on the properties of categories) as well as more simple models. In simple models, categorisation results from the experience on the stimuli (prototypical or exemplar-based templates) (Ashby & Maddox, 2005).

One approach in studying the phonetic representation of vowels is to consider their role as the target of phonetic acts (Eerola, Laaksonen, Savela & Aaltonen, 2003a; Eerola, Laaksonen, Savela & Aaltonen, 2003b; Eerola, Savela, Laaksonen & Aaltonen, 2003; Alivuotila, Hakokari, Savela, Happonen & Aaltonen, 2007; Alivuotila, Savela & Aaltonen, 2008). In Guenther's DIVA model (Guenther, Hampson & Johnson, 1998), the motor control of the speech production/perception system is considered to be a type of planning framework of gestural acts. Sounds are represented as targets that are auditory regions in perceptual vowel space, including acoustic information on the vowel in question. As the same vowel can be produced using a variety of gestural strategies, the model emphasises the auditory targets instead of articulatory gestures. According to Guenther, an oro-sensory control of speech targeting exists, but it is less effective for vowels than for consonants. Guenther also stresses that auditory properties are more constant targets of articulation than co-articulated gestures. While the auditorist-gesturalist debate remains open, auditory feedback, at least in the case of vowels seems to be more direct than gestural (see, for example, Gay et al., 1981).

Prototype-based learning models have signified the emergence of sound categories in the perceptual vowel space of children and adults (Kuhl, 1991, Kuhl, 1993). In the magnet theory, a higher rating in the prototypicality (goodness) of vowels is associated with lower sensitivity in vowel discrimination. In their crucial study, Iverson & Kuhl, (2000) established that the magnet effect is not related to the global identification of sounds and that, while the goodness rating was not affected by the nature of preceding vowels, the identification was. This was interpreted to mean that the magnet effect stemmed from an earlier process rather than a typical contrast effect, which is often manifested in the identification functions of categories. Using the terminology of this thesis, the symbolic level of identification was affected by the context effect but the indexical level was not.

2.3.4. Euclidean formant-based vowel space and phonetic similarity

The Euclidean distance is a general model of the psychometric spaces in psychology (Shepard, 1974). The view that formants are the primary carrier of the phonetic quality of vowels has received direct support from experiments in phonetic distance estimations. Studies on the evaluation of spectral differences have shown that listeners favour formants as important features in similarity estimations of spectra (Chiba & Kajiyama, 1958; Pols, Van der Kamp & Plomp, 1969; Carlson & Granström, 1979). The Euclidean distance in space defined by F1 and F2 has been favoured as the basis for the perceptual distance between vowel stimuli. Assman and Summerfield (1989) demonstrated that, if the spectral slopes between formant peaks in a particular vowel category are changed to a flat background noise and the formant peaks are changed to sine waves that represent the formant peaks, the manipulated

stimuli can still be identified on the basis of peak frequency values of F1 and F2. They concluded that the spectral slope (e.g. spectral moments) cannot explain the vowel similarity in this type of experiment because different vowels have different slopes that disappear with a change in the amplitude pattern of a particular sound type.

2.3.5. Asymmetries and the Euclidean vowel space on phonetic similarity

Formant values have been challenged as a global indicator of vowel quality. In a study by Bernstein (1981) it was found that, in addition to formant frequencies, formant bandwidths affected the evaluation of vowel similarities. On the other hand, there have been suggestions regarding local fusions of formants in the perceptual vowel spectrum (see, for example, Bladon & Lindblom, 1981; Chistovich & Chernova, 1986). Direct comparisons between formants and the whole spectral shape in the perception of phonetic quality have also been made. In Klatt's classic study (1982), subjects were asked to determine the phonetic quality of the stimulus. Small shifts in formant frequencies affected the phonetic quality, whereas the similar shifts in the spectral slope did not.

2.3.6. Alternative models on vowel similarity: the role of tilt and amplitude ratio

Two recent studies have challenged the notion of formant-based vowel identification. In a study by Ito et al. (2001) the suppression of the peak in the F1 area did not have an impact on the identification of Japanese vowels, as long as the amplitude ratio between lower and higher frequencies (the crucial frequency being 1500 Hz) was not manipulated within the stimulus set. In a study by Kiefte and Kluender (2005), the pattern recognition of vowel spectra was studied by manipulating spectral tilts (that is, a spectral balance measure). Identification of synthetic stimuli was shown to be affected both by spectral tilt and by the similarity of formant frequencies. A seven-step continuum from [i] to [u] was used, and some of the vowels had the same F1 and F2 values but different slopes in amplitudes. The results showed that the change in tilt did affect identification of the stimuli although the tilt effect was less than the effect of the two lowest formants. Furthermore, strong tilt in spectrum seemed to favour the [u]-responses.

In conclusion, while the formants can be expected to be the primary indicator of vowel quality, additional attributes can affect phonetic categorisation. The formant space and the space with both formants and spectral moments are contrasted in this thesis, for discussion e.g. (Savela & Pikkanen, 2005). The identification response and the discrimination response are compared in different combinations of F1 and F2 values, which reveals the acoustic features

that determine the differences between sounds and the features used to name the sound on the basis of the symbolic task (identification).

2.3.7. Asymmetry of the vowel prototype distribution

The distribution of vowel prototypes has seldom been investigated on the level of the perceptual vowel space. Although the original theory on perceptual magnets expected the prototypes to be in the middle of the categories (Kuhl, 1991), it was shown later that identification boundaries and location of prototypes were independent of one another (Iverson & Kuhl, 2000). On the other hand, Lacerda (1995) presented an exemplar-based model in which identification boundaries were based on the distribution of vowel exemplars within the language input. However, no comprehensive data has yet been presented at the language level and in the terms of classical phonetic features.

The assumption that vowel perception reflects index-based (i.e. prototype based) learning of sound similarities has not met with unanimous support in the relevant literature (for example, see Polka (2003)). In the study by Aaltonen et al. (1997) some subjects, good categorisers, rated vowels with the highest grades near the category boundary, while some subjects who were poor categorisers rated vowels with the highest grades in the extreme positions in the vowel continuum, far from the category boundary. In contrast to the poor categorisers, the good ones had a steep category structure and, in discrimination tasks, had significant changes in MMN-responses in relation to the prototypicality of the stimuli. In later studies, discrimination has not reflected the magnet effect in which goodness equates to poor discrimination ability. The prototypes were located in peripheral locations of vowel space by all subjects, (e.g. Sharma & Dorman, 1998; Thyer, Hickson & Dodd, 2000).

2.3.8. Phonemic vowel theories and pattern recognition studies of vowels

There are several theories regarding physical attributes linked to the phonemic classification of sounds. Two models, feature-based models and pattern recognition models, have been used. In the feature-based models, sounds are perceived in accordance with the distinctive features (e.g. Jakobson, Fant et al., 1961). These features can be acoustic or articulatory, but they are based on implicit knowledge regarding their perceptual/symbolic value.

Further studies on speech perception have explained linear regression models on the identification of speech sounds by selecting some psycho-acoustic dimensions in physical space and analysing how these factors follow the identification. This is called a pattern recognition model of speech perception. The experimental procedure in these studies is usually based on a limited set of symbolic categories. Nearey (1989) explored the effects of certain suggested characteristics on symbolic vowel identification. Dimensions of special interest have been so-called *intrinsic* and *extrinsic* factors in vowel perception. The intrinsic factors are the primary features of vowels, for

example, formants (F1 and F2) and fundamental frequency (F0). In extreme intrinsic theories, vowel identification is based on certain relationships between the formants and the fundamental frequency (Miller, 1989). Instead, the extrinsic factors, which are based on context and speaker-dependent properties, are not always found in single-vowel segments but can also be detected in longer utterances. In fact, according to this theory, the identity of a vowel can be recognised mainly in the context of other sounds. The so-called dynamic model of vowel perception suggests that, to a large extent, the quality of the vowels is determined by the context (Strange, 1989).

Nearey and Kiefte (2003) compared the strength of different auditory representations in Finnish and English. The vowel categories were compared with different models in which the effects of higher formants are integrated, for example the F1F2F3 model, the F1F2 ' model and the PLP (perceptual linear predictive) model (Hermansky, 1990). The PLP was proven to be the best model and the F1F2 'the poorest one.

This thesis examines the relationship of indexical (experience-based) and symbolic (phonological) aspects of vowel perception. This is done by comparing the indexical aspects (indicated by the goodness rating) and the symbolic aspects (indicated by identification data) in terms of their different acoustic features. This thesis will argue that these levels differ in the use of acoustic attributes. The role of biology is studied in the context of indexicality and the role of language is studied in the context of symbolicity.

3. STUDY I: IDENTIFICATION OF ISOLATED SYNTHETIC STEADY-STATE VOWELS

Study I examines the categorisation of vowels by using a set of synthetic vowel stimuli. The identification and goodness rating responses are compared in different languages by studying how the acoustic attributes are reflected in the vowel charts of those languages. The acoustic attributes are examined in three analyses. The identification of Udmurt vowels is investigated in *Analysis 1*. Udmurt was chosen due to its non-continuous nature of F2 in the vowel space. This phenomenon was observed in the preliminary analysis of TVT.

The identification responses for four languages – Finnish, German, Spanish and Czech – were compared in *Analysis 2*. Finnish and German each have eight vowel categories but slightly different phonetic descriptions. Czech and Spanish each have five vowel categories but slightly different phonetic descriptions. The prototypical vowels of these languages were compared in *Analysis 3*.

The data clarifies the acoustic criteria affecting the categorisation of vowels and therefore provides a tool for investigating the differences in the categorisation process of different languages. Furthermore, Study I explores the problem of vowel categories as interpretants, that is, whether they reflect the iconic, indexical or symbolic categorisation principles.

3.1. Vowel identification experiment

This chapter presents the structure and methodology of the TVT study used in this thesis.

3.1.1. Methods

Vowel systems of the studied languages

The conventional phonetic descriptions of vowel classes, described with the IPA alphabet (IPA, 1949), are presented in Figure and are used throughout this thesis. The IPA symbols are used according to phonetic conventions and are based on systemising the impressions of the phonetic features according to the salient attributes of the sound. They are based on both the absolute acoustic value and on the relative properties (so-called shape) of sound systems. The shape refers to a theoretical model on the distinctions that are important to particular languages. The model can be based on a phonological, phonetic or acoustic analysis of sound systems and, therefore, phonetic descriptions can use the same symbols inconsistently.

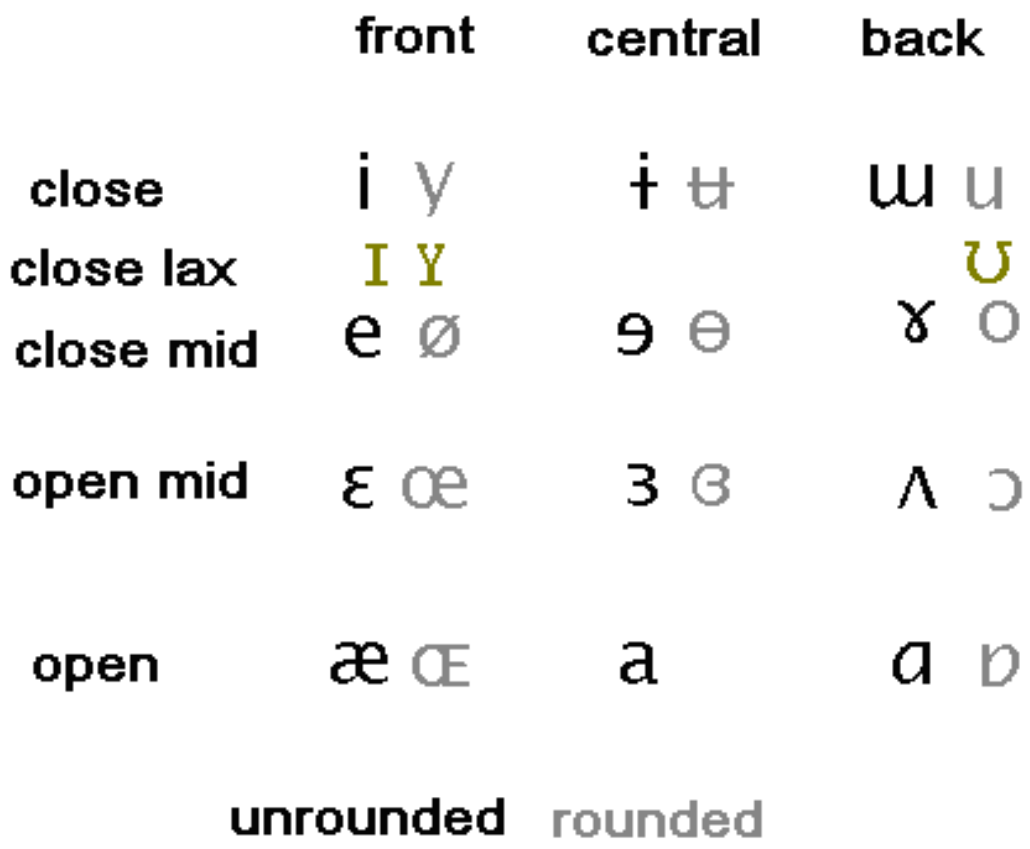


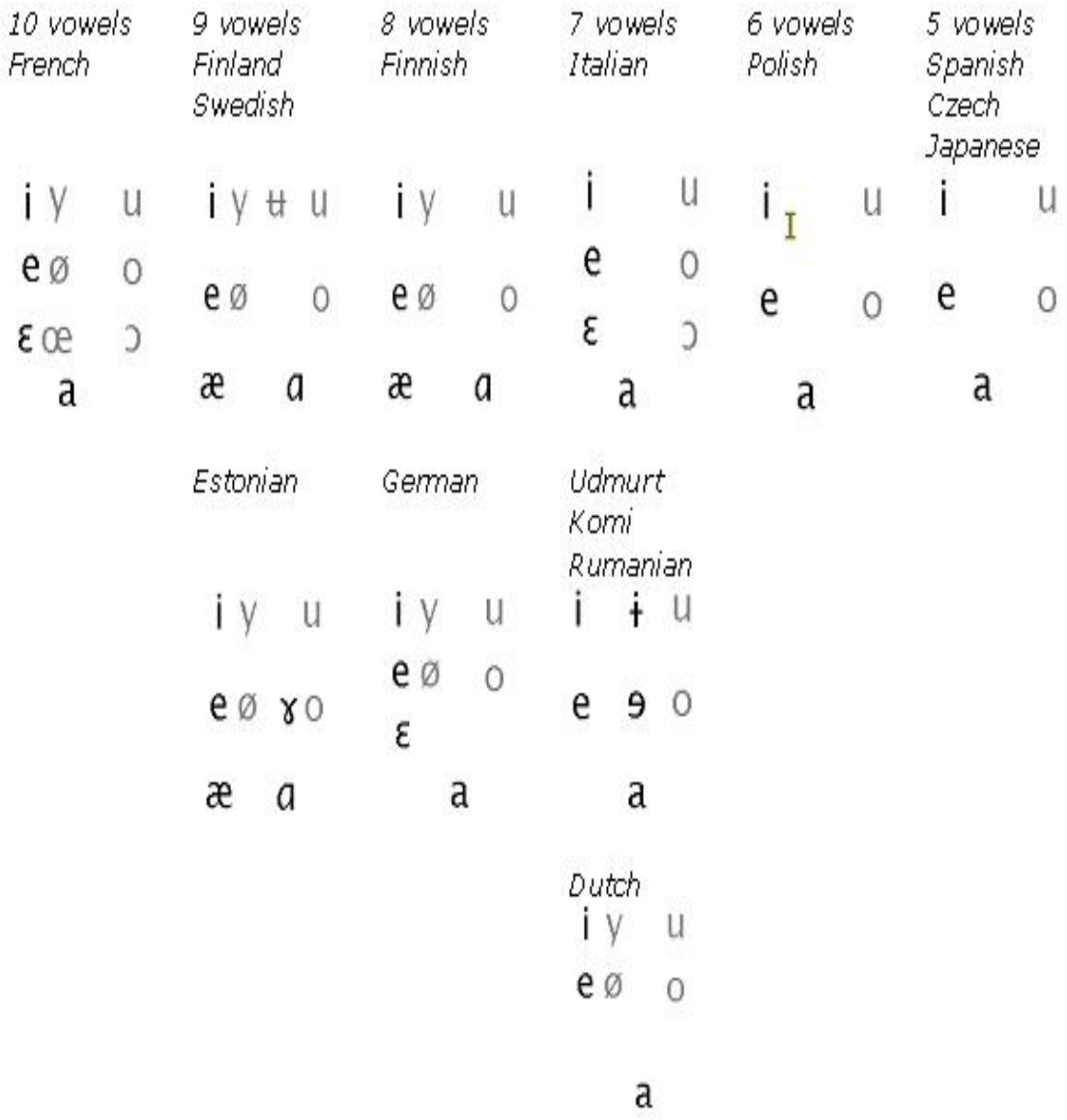
Figure 2. Phonetic symbols used in this thesis
Phonetic symbols are classified according to their conventional phonetic names (IPA, 1949). The symbols of the close mid-vowels are used also for the mid-vowels if the language has only three different classes in vowel closeness. Where symbols appear in pairs, the one to the right represents the rounded vowel.

Study I concerns the relationship between the acoustic, phonetic, and phonological levels of information in 14 languages (Table 1).

Stimuli

The test consisted of synthetic vowels that covered the entire vowel space except for diphthongs and nasal vowels (Figure 3). The stimuli were synthesised using a Klatt serial synthesiser (1980). The vowel space was created by varying the peak frequency of F1 from 250 to 800 Hz with steps of 30 mels and F2 from 600 to 2800 Hz with steps of 50 mels. The peak frequency of F3 is 2500 Hz as long as F2 is 2000 Hz or below and is higher by

Table 1. Vowel systems of the languages studied based on the definitions of the UPSID data base (Maddieson & Disner, 1984)



200 mels when F2 is above 2000 Hz. The duration of the vowel stimuli was 350 ms and their fundamental frequency firstly rose from 100 Hz to 120 Hz (until 120 ms) and then fell to 80 Hz during the rest of the stimulus.

In the TVT data set, the changing parameters were the lowest formant values. According to Klatt, the amplitudes and bandwidths of the formants change according to the set of rules. The principal rule is that the amplitudes of higher formants are affected by manipulating the first formant. The KLATT synthesiser

is based on a one-dimensional wave equation that can be considered valid below 5 kHz. The system models a vocal tract that is 17 cm with five resonators, and the amplitudes and bandwidths of formants are given by formulae that are intended to resemble the human voice. The amplitude of the formant peak is inversely proportional to its bandwidth. If the formant bandwidth is doubled, the vowel amplitude is increased with 6 dB. For example, if the formant bandwidth is halved, the peak amplitude is decreased by 6 dB. The formant amplitude is also dependent on the formant frequency. If the frequency of the formant peak is doubled, its amplitude is increased by 6 dB. The amplitudes of the higher frequencies are dependent on the amplitude of the lower peaks by a factor proportional to the frequency squared. For example, if the frequency of the lower formant is halved, the amplitudes of higher formants are reduced by 12 dB.

Subjects

The subjects were made up of students, exchange students, and personnel at the University of Turku (Table 2).

Table 2. The number of subjects in different language sets

Language	*Number of subjects*	*Mean age of subjects*
French	*3*	*25.6*
Swedish in Finland	*19*	*34.6*
Estonian	*4*	*30.0*
German	*19*	*25.3*
Finnish	*68*	*26.2*
Dutch	*8*	*24.2*
Italian	*9*	*27.5*
Udmurt	*6*	*28.5*
Romanian	*3*	*35.0*
Komi	*2*	*28.3*
Polish	*5*	*41.0*
Spanish	*17*	*24.5*
Czech	*10*	*27.1*

3.1.2. The Turku Vowel Test

The Turku Vowel Test, known as TVT (the laboratory model) was constructed using an application used in the language teaching laboratory of the University of Turku. The stimuli were presented via earphones on a personal computer or via loudspeaker if the test took place in an individual room.

In the test the subjects were asked to choose their language from the alternatives presented to them. Subjects chose the vowel category from the options presented on the screen and, by using the mouse button, made a goodness rating. They rated the goodness of stimulus (scale: 1-7) as a

member of the chosen category. It was also possible to repeat the current stimulus or pause the test and continue later.

Vowel data from the test was used later in different types of analysis. The test equipment itself did not include a chart plotter, as a different JAVA applet was used for that purpose. The pictures shown in Study I were made using a programme developed for that purpose.

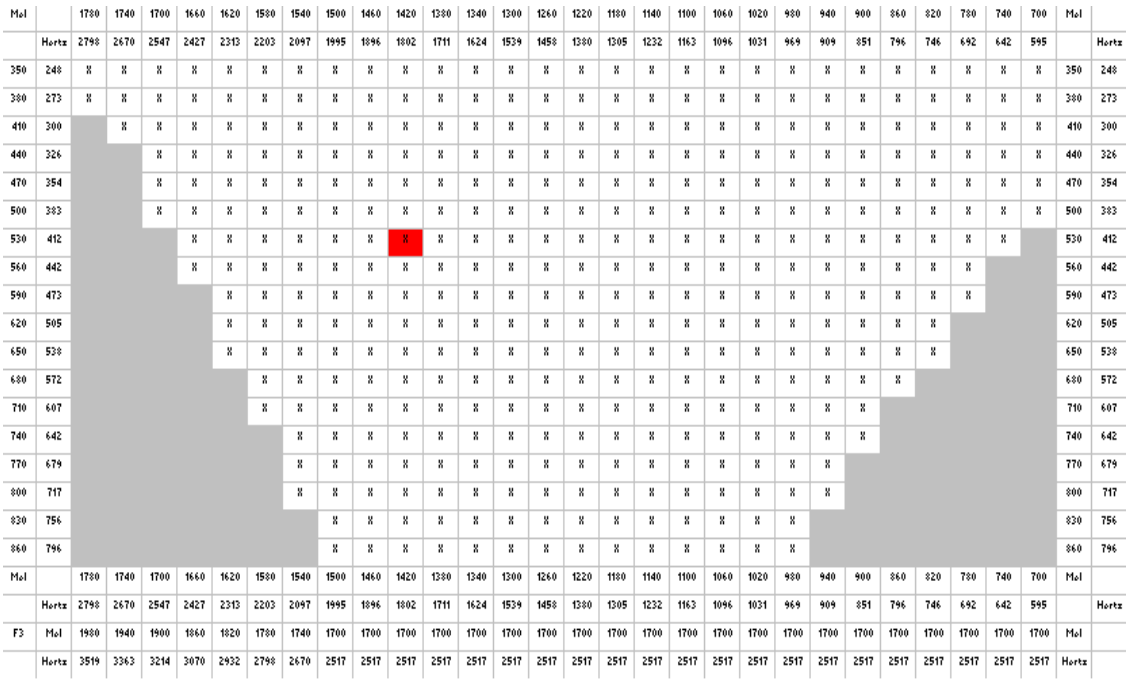


Figure 3. Vowel plane in the identification experiment
The Mel and Hz values of the stimuli are plotted in the diagram. The peak frequencies of the F1 peak frequency values are in the horizontal axis and F2 and the peak frequency values of F3 are plotted on the vertical axis. Corresponding mel and Hertz values are indicated; for example, the sound coloured has F1 412 Hertz, F2 1802 Hertz and F3 2517 Hertz.

3.2. Descriptive analysis of the data from the Turku Vowel Test

The hypothesis of Study I was that, in vowel identification, the knowledge on prototypicality (indexicality) and vowel boundaries (symbolicity) is based on the varying use of spectral attributes (formants as well as spectral moments). The experiments provide information on the symbolic categorisation of the stimuli. The structures on the vowel chart are understood to reflect the iconic and indexical modes of vowel perception. Finally, the question of whether the results show a real symbolic system is a valid one. The locations of all the other vowels affect the identity of the vowel stimulus so that the boundaries are half way between the two prototypes.

The iconic mode means that easily detectable and discernable features are used in the identification of vowels. Typical theories that use the iconic mode as a main explanation of sound systems are different distinctive features theories such as the quantal theory of Stevens (1989) or the auditory enhancement theory of speech (Kingston & Diehl, 1994). These theories are, however, based on generative theory linguistic knowledge and are therefore difficult to interpret in this study, which is based on the pattern recognition tradition of vowel studies.

In this thesis 'terminology, the indexical mode upon which the categorisation is based on is prototypes, and the identification is based on matching the input to the linearly closest prototype in the formant space (e.g. Rosner & Pickering, 1994).

Finally, according to the terminology of the thesis, the symbolic level/mode expects that the identities are based on the distinctions between existing categories of the vowel system, not the absolute physical values of the sounds. (For example, in Finnish the vowels of the [a] -area are placed either into the /æ/ or into the /ɑ/ –category, since the /a/ -category does not exist in the Finnish vowel system). This thesis argues that additional iconic information about the signal (e.g. spectral moments) could also be used in identification, if the perceived stimuli are non-prototypical.

The identification and goodness rating criteria the vowel categorisation according to their phonological systems are studied from simpler to more complicated ones. The first cases under investigation are the languages with five-vowel systems (Figure 4).

Czech and Spanish differ in their pattern of prototypes. The Czech /e/ is often considered to be more open than the Spanish /e/ (Maddieson & Disner, 1984), and in Czech the /e/ -category is wider than in Spanish. In Czech there is also a large region of non-prototypical /i/ vowels. A similar region is also found in Polish, a language that is historically related to Czech, even though the region is divided into two separate categories (see Figure 5). Polish results for identification functions have also been presented by Jassem (1968).

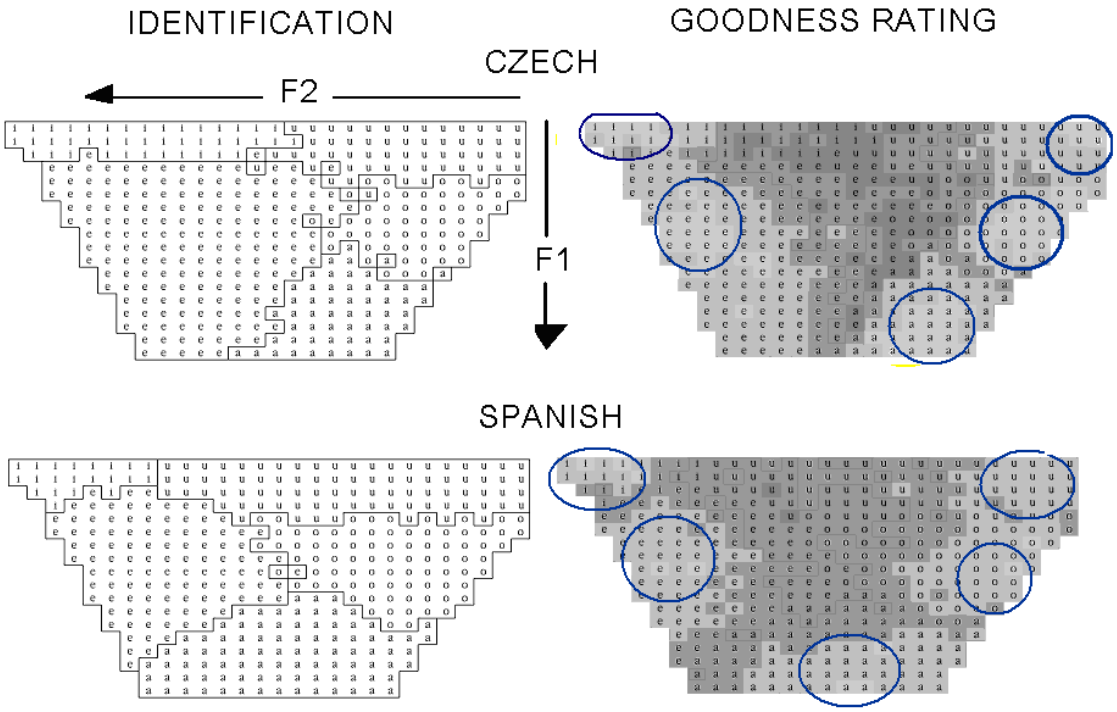


Figure 4. Czech (upper) and Spanish (lower) vowel charts
The formant values are described in Figure 3. In the goodness rating charts, lighter areas represent higher goodness ratings. Circles represent the areas with the highest ratings.

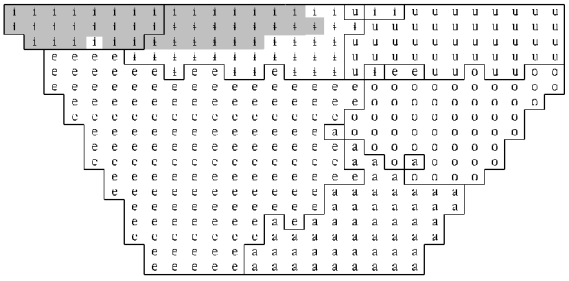


Figure 5. Polish vowel categorisation (Czech /i/ shaded)

Estonian and Finland-Swedish both have nine vowel systems (see Figure 6). The systems differ in that, in Estonian the close central vowel is an unrounded /ɤ/, whereas in Finland-Swedish it is a rounded /ʉ/.

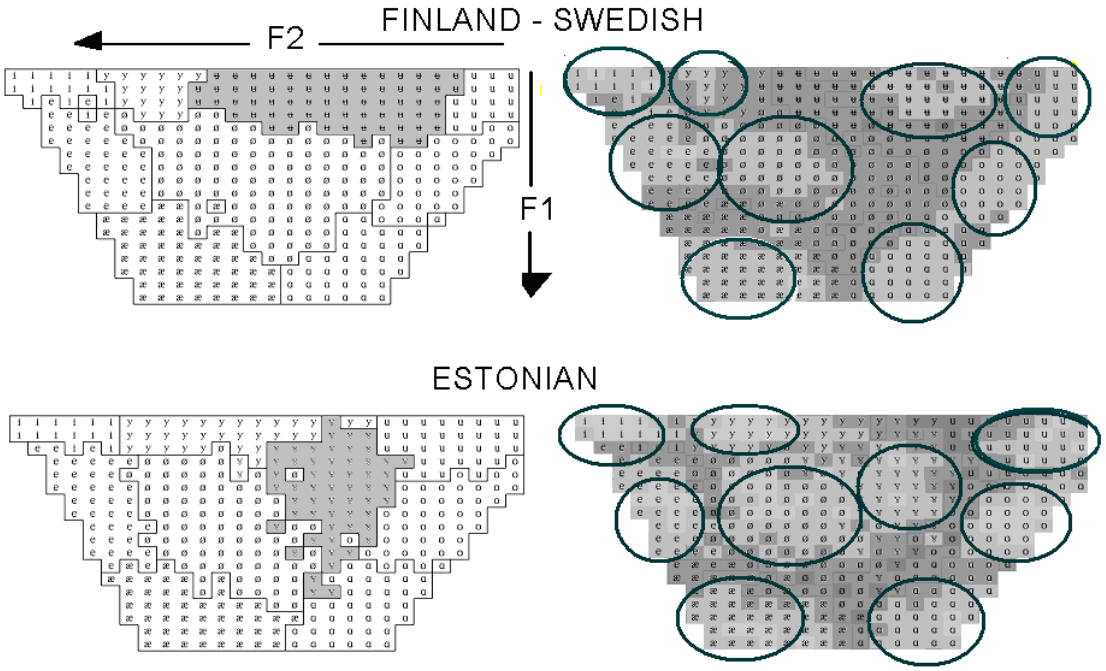


Figure 6. Estonian and Finland-Swedish vowel charts
The formant values are described in Figure . The vowel between /y/ and /u/ is shaded. In the goodness rating charts, lighter areas represent higher goodness ratings. Circles represent the areas with the highest ratings.

The vowel systems of Estonian and Finland-Swedish differ in terms of prototypes, and the vowel boundaries are not affected by the distance between prototypical vowels, as can be seen near the /o/ -area. The areas are similar despite the more closed Estonian vowel /ɤ/ compared to the Finland-Swedish /ʉ/. The Finland Swedish results follow the patterns found by Määttä (1983).

The next two charts present vowel systems with more than three front and back vowels. In French and Italian the vowels /e/ and /ɛ/ are relatively similar in F1 but differ in F2. A similar variance can be seen more evidently in the back vowels /o/ and /ɔ/. The French /ɔ/ does not reach as far back as the Italian /ɔ/.

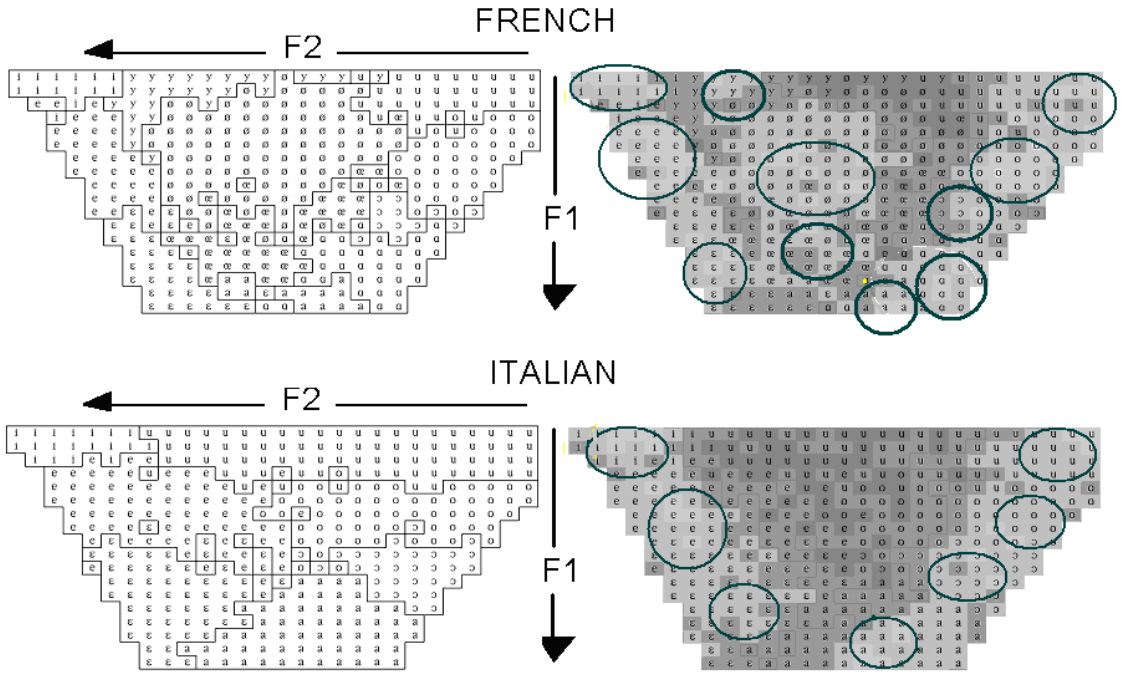


Figure 7. French and Italian vowel charts
The formant values are described in Figure . In the goodness rating charts, the lighter areas have higher goodness ratings. Circles represent the areas with the highest ratings.

Additional determinants of boundaries and prototypes: spectral moments

Spectral moments can be used as an additional feature in vowel perception. The measurement of spectral moments is based on the power spectrum in which the magnitudes of spectral components are squared. It has been used by Forrest (1988) in the classification of fricatives and by Milenkovic et al. (1988) to classify vowels. The basic problem with spectral moments is that they merge individual voice differences with possible speaker-independent phonetic aspects (such as the filter function of the vocal tract). There can also be pathological differences between individual speakers which are reflected in spectral moments (Flipsen, Shriberg, Weismer, Karlsson & McSweeney, 1999).

The spectral moments are whole-spectral features that describe the distribution of spectral energy within the frequency range of a vowel. Vowel similarities that may be based on the use of spectral moments are described in Figure 8. They may relate primarily to vowel systems with central vowels.

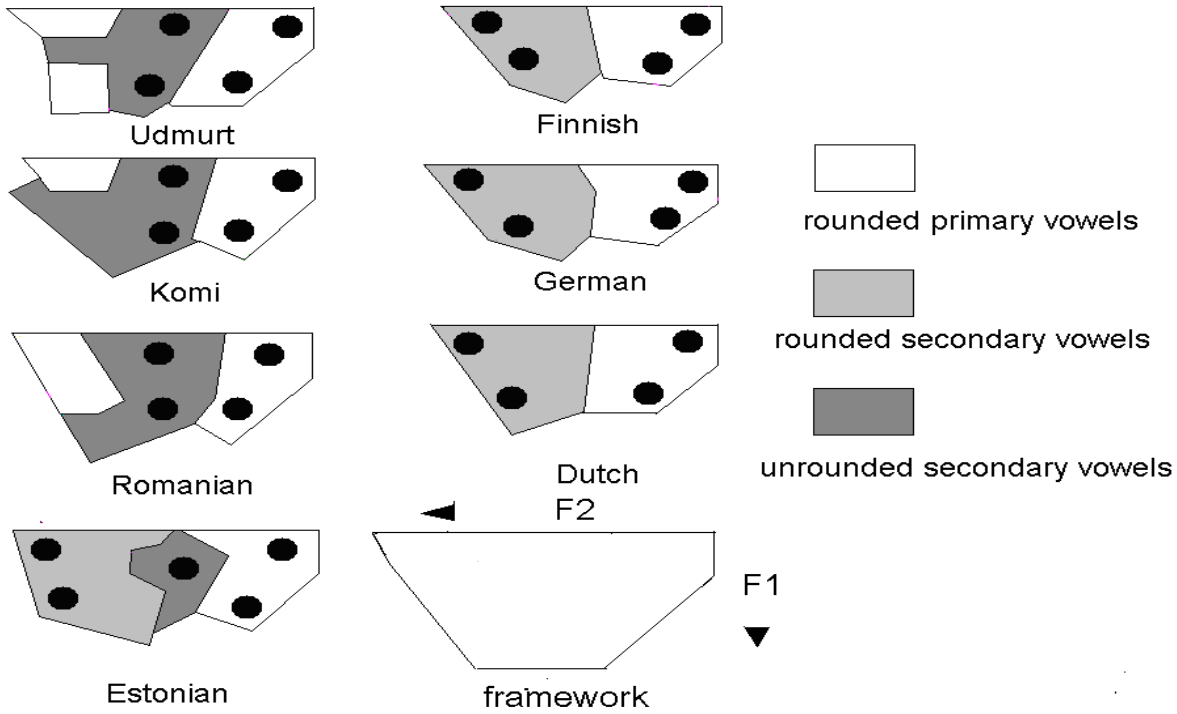


Figure 8. Vowel systems with secondary vowels
Vowel systems in which the identification of secondary vowels may be based more on other criteria than the formants. Primary vowels are the most common vowel categories in languages, whereas secondary vowels (e.g. non-open unrounded back vowels and rounded front vowels) are considered less common.

There are four different spectral moments: centre of gravity, standard deviation, skewness and kurtosis, which will be discussed in detail next. In general the *n*th central moments of the spectrum can be calculated for complex stimulus S (f) by $\int_0^\infty (f - f_c)^n \, |S(f)|^p \, df$ divided by $\int_0^\infty |S(f)|^p \, df$.
The second, third and fourth central moments are referred to as variance, non-normalised skewness and non-normalised kurtosis. Spectral moments are measured in this study after FFT transformation of the sound spectrum throughout the duration of the whole stimulus. The analysis was made using PRAAT analysis equipment (version 4.5.18) (Boersma & Weenink, 2001). Five thousand five hundred samples were used and the sampling frequency for the stimuli was 11000 Hz.

Centre of gravity (CoG)

The CoG represents the average of the frequencies across the entire spectrum (Figure 9). For example, a sinusoidal stimulus of 500 Hz has a CoG of 500 Hz. For white noise, the CoG is half of the *Nyquist frequency* of the signal, while the Nyquist frequency, in turn, is half the sampling frequency of the signal. The

CoG for the sine wave is equal to the wave's frequency (Oppenheim, Schafer & Buck, 1999). For more complex sounds, like vowels, different distributions of the spectral amplitude may have the same CoG. In Praat, the CoG for a complex stimulus S (f) is given by $\int_0^\infty f\,|S(f)|^2\,df$ divided by "energy" $\int_0^\infty |S(f)|^2\,df$.

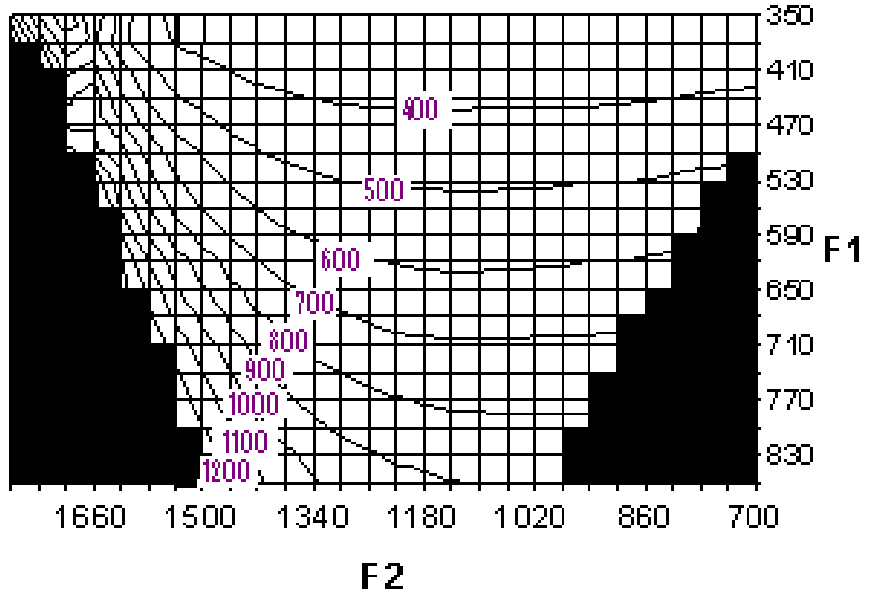


Figure 9. Centre of gravity in a power spectrum of synthetic vowels plotted against formants (in a mel scale)
The lines present the frequency contours of the CoG (in a mel scale).

The CoG continua reflect the identification of secondary front vowels (/ø/ and /y/) in some languages (Figure 10). In Dutch, German and Finnish languages, the boundaries are more similar in rounded vowels than they are for non-rounded ones. The Dutch /ø/ is closer to the German /ø/ than to Finnish /ø/. However, the most distinctive features are the contrasts between the lower open front vowels, /e/, /ɛ/, and /æ/. The prototypical /e/ seems to be closer (smaller F2) in the Germanic languages than in Finnish, but the boundaries are different in shape.

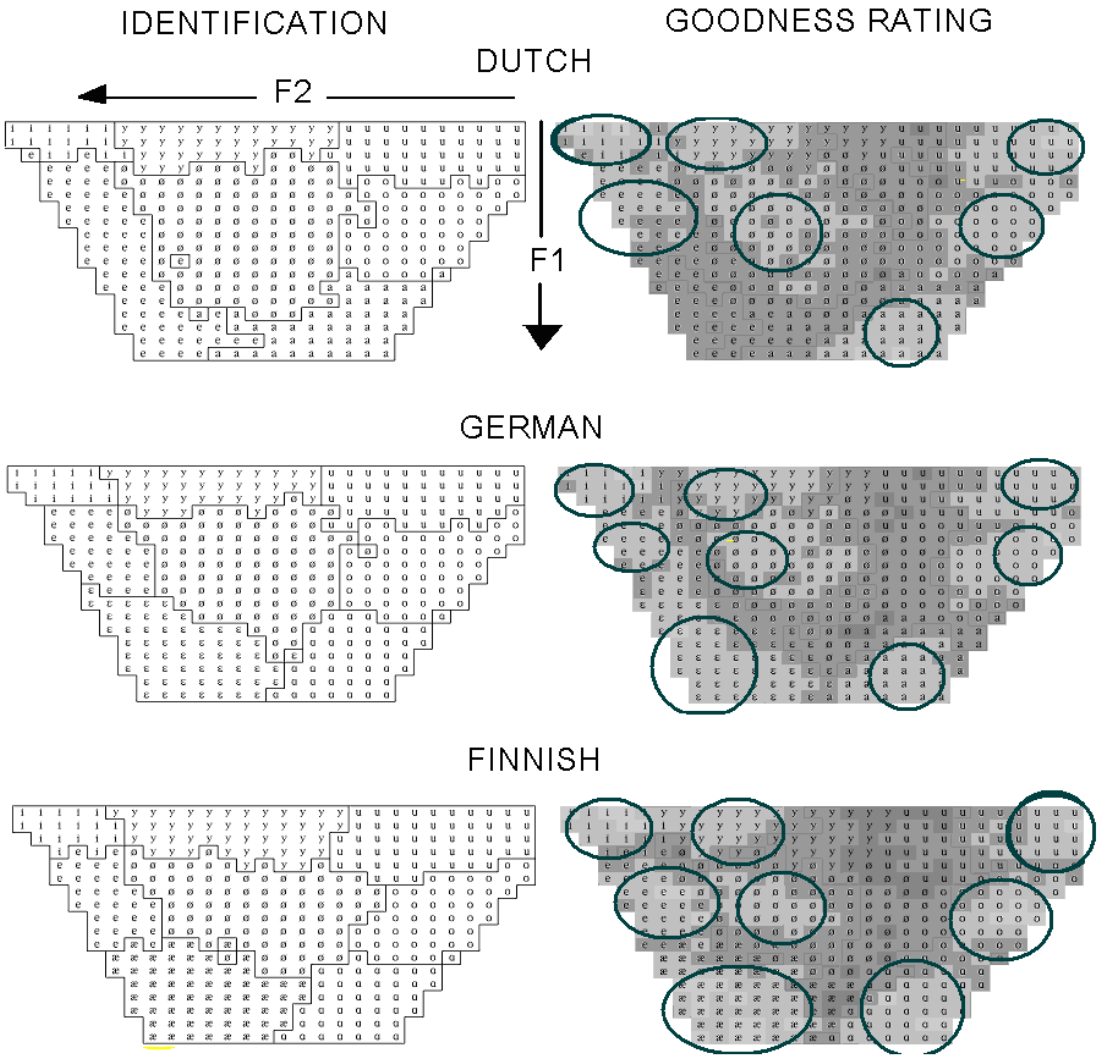


Figure 10. Dutch, German, and Finnish vowel charts
In the goodness rating charts, the lighter areas have higher goodness ratings. The circles represent the areas with the highest ratings.

Both the boundaries and the set of vowels considered to be good vowels seem to reflect the CoGs. In the vowels with high F2 and F3, the effect of CoG is similar to F2 ,' which is an integrated formant on the basis of higher formants (Carlson, Granström & Fant, 1970; Schwartz, Boe, Vallee & Abry, 1997a). In Japanese the identification criteria reflects the variance of CoG (Figure 11), while the goodness rating seems to be less dependent on CoG. Note the boundary between the rounded back vowels (/o/ /ɯ/) and the other vowel categories. However, the good /o/ are located mostly on low F2 area of the vowel space, whereas the prototypical /ɯ/ is more widely spread out on the F2 axis.

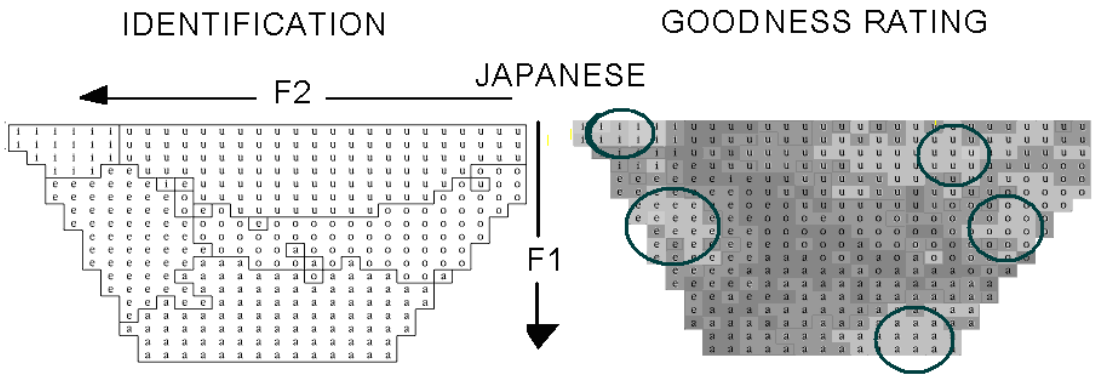


Figure 11. Japanese vowel chart
In the goodness rating charts, the lighter areas have higher goodness ratings. The circles represent the areas with the highest ratings.
Standard deviation (Std)

The standard deviation describes the extent to which the frequencies of the spectrum deviate from the CoG, on average. Std is the square root of the second central moment of the spectrum (Figure 12).

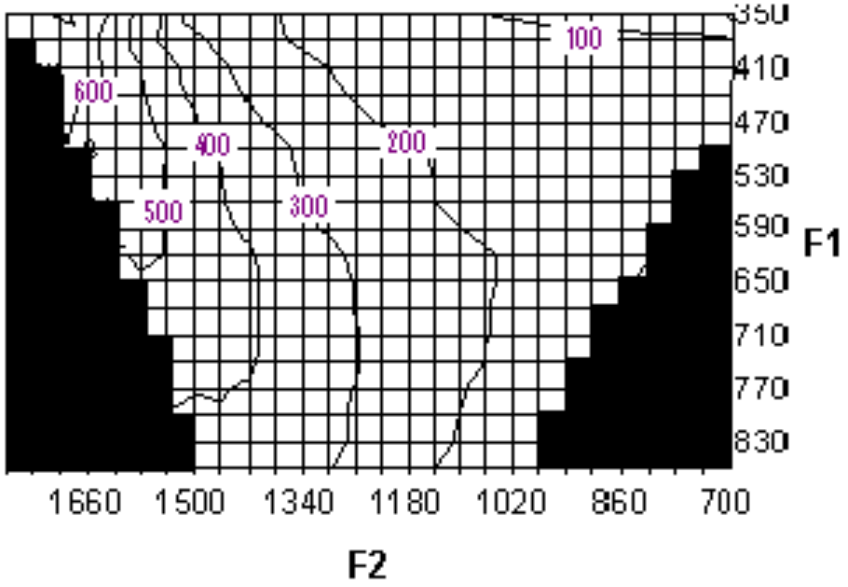


Figure 12. Standard deviation in a power spectrum of synthetic vowels plotted against formants (in a mel scale)
The frequency contours of the Std.

The standard deviation may be a secondary feature that reflects the identification of the front–back dimension in Czech (especially in /e/) (Figure 13). All /e/ answers seem to have high F2, although that does not necessarily hold /i/ answers.

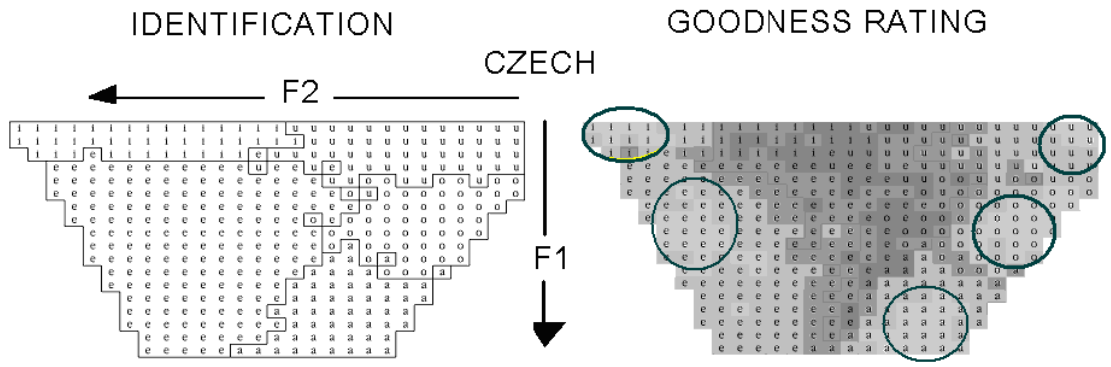


Figure 13. Czech vowel chart
In the goodness rating charts, the lighter areas have higher goodness ratings. The circles represent the areas with the highest ratings.

The same relationship can be seen in Figure 14.

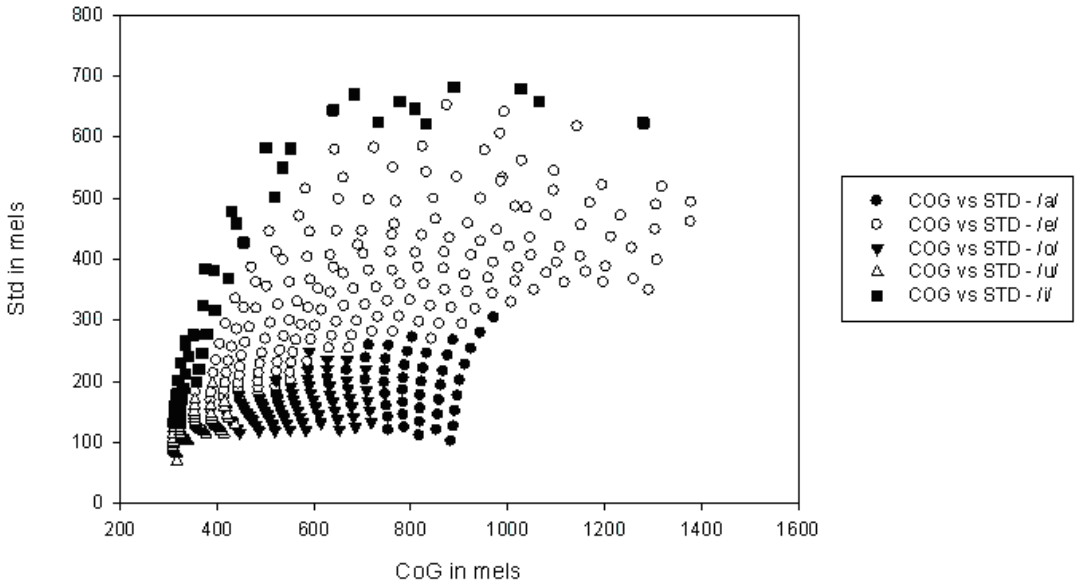


Figure 14. Czech vowel answers (the most frequent answer for a particular sound stimulus) plotted against CoG and Std

Skewness

The (normalised) skewness presents the asymmetry in the shape of the spectrum between the frequencies above and below the CoG. It is measured by dividing the (non-normalised) skewness by the 1.5th power of the second spectral moment (Figure 15), and. It can be presented with the formula

$$\frac{\mu_3}{\mu_2^{3/2}}_i$$

in which μ_2 is the second spectral moment and μ_3 is the third spectral moment.

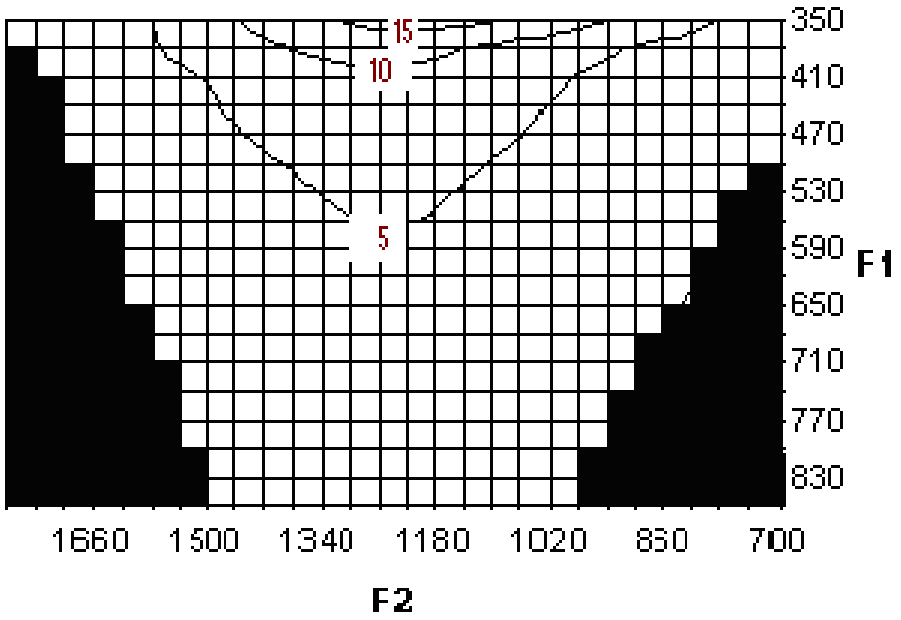


Figure 15. skewness in a power spectrum of synthetic vowels plotted against formants (in a mel scale)

Few languages have primary vowel categories in the area with high skewness, so vowels with high skewness seem to be avoided. In central vowels, however, skewness may be used as a prototypicality criterion.

Kurtosis of the spectrum

The (normalised) kurtosis of the spectrum describes the extent to which the spectrum differs from the Gaussian distribution of the spectrum (Figure 16). This is measured by dividing the non-normalised kurtosis by the square of the standard deviation, and subtracting 3. The resulting value indicates the extent to which the distribution of spectral components around the CoG differs from the Gaussian shape for the FFT spectrum.

It can be represented by formula

$$\frac{\mu_4}{\mu_2^2} - 3$$

in which μ_2 is the second spectral moment and μ_4 is the fourth spectral moment.

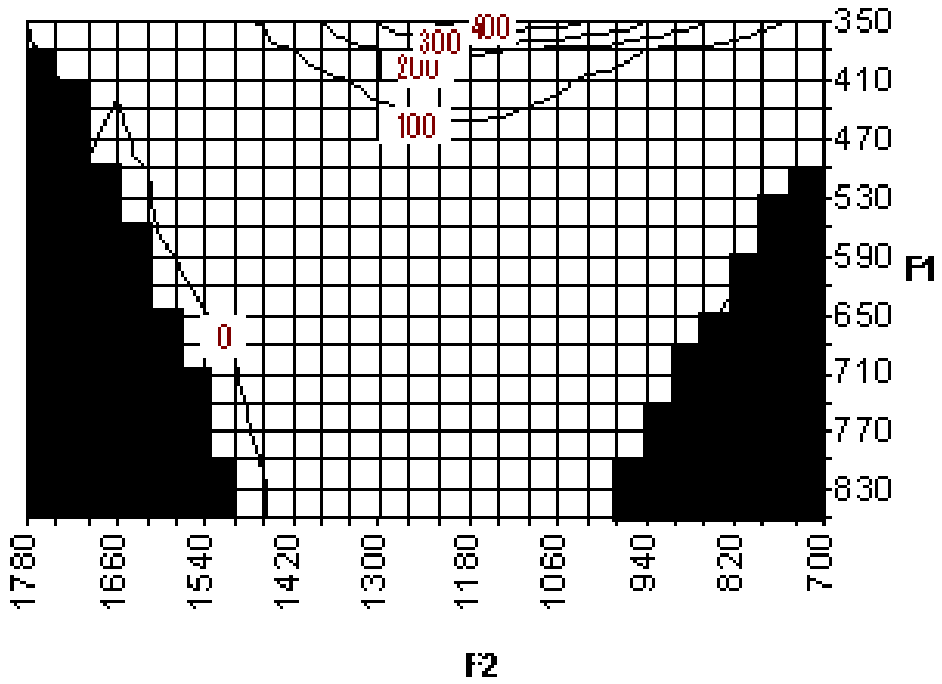


Figure 16. Kurtosis in a power spectrum of synthetic vowels plotted against the formants (in a mel scale)

The pattern of kurtosis reflects the pattern of skewness, especially in areas with low F2 (around 1100 mels), and this may affect the identification of unrounded back vowels. In Japanese, the only language in which the prototypicality of a category seems to reflect the kurtosis, the prototypical /u/ has a higher F2 than the /o/ (see Figure 17). The non-continuity of Romanian, Komi and Udmurt /u/ responses may also be based on kurtosis and skewness (Figure 17).

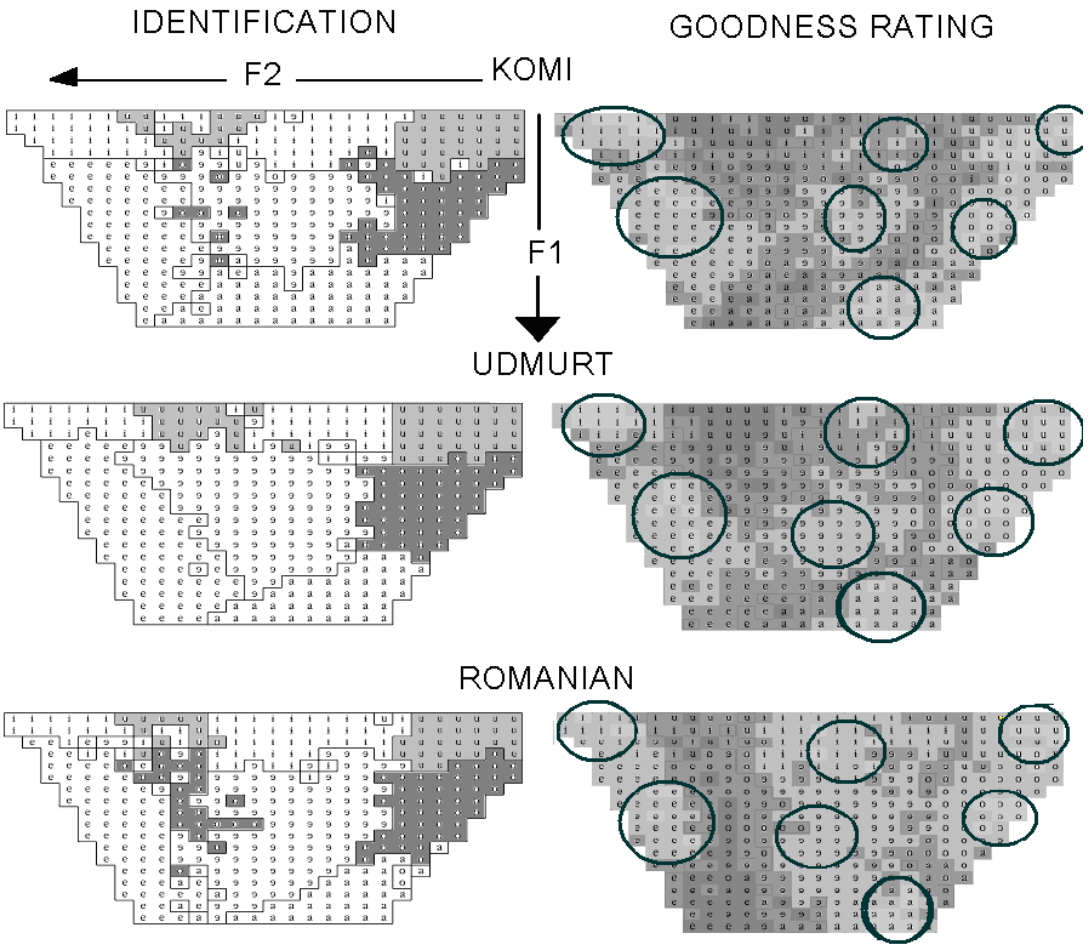


Figure 17. Komi, Udmurt and Romanian vowel charts
In the identification charts the /u/ and /o/ -responses are shaded to emphasise their non-continuity.

The problem with using spectral moments in vowel identification becomes more complicated when the criteria for individual vowel classes are investigated. For example, kurtosis and skewness may explain why the prototypical vowels are situated peripherally in the vowel space (Figure 18). A variety of combinations of secondary acoustic attributes are also possible, for example, Kurtosis (over 200) and CoG (at 350 mels) can be used in the identification of /u/ in Udmurt.

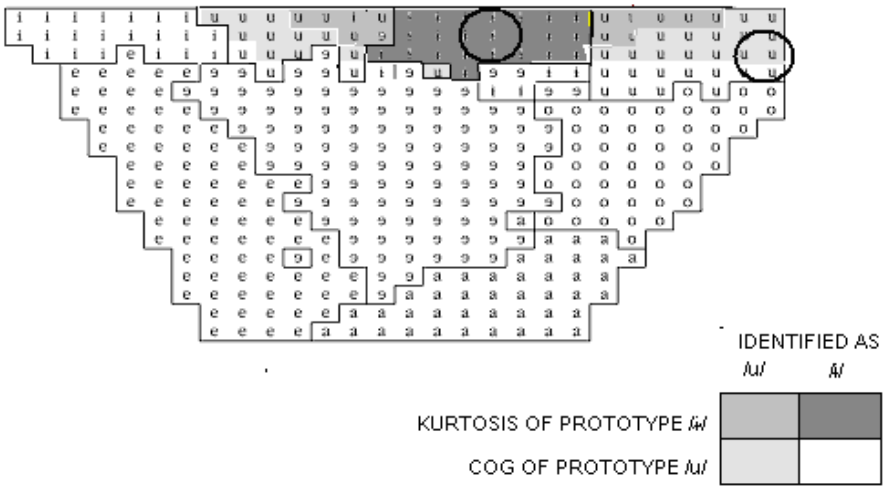


Figure 18. Udmurt vowel chart – the use of additional spectral attributes

The circles define the areas of the prototypical /u/ and /ɨ/. The lightly shaded area has the same CoG as the prototypical /u/ and is identified as /u/. The next lightest areas represent the areas in which the /u/ responses have the same Kurtosis value as the prototype /ɨ/. The darkest area is identified as /ɨ/ and has the same Kurtosis value as the prototype /ɨ/.

The previous figure (Figure 18) can be redrawn as in the next figure. The areas of vowel categories are more difficult to see than in previous studies. The following figures (Figure 19 and Figure 20) demonstrate that, in general, the non-rounded central vowels have higher skewness and kurtosis than other categories among the majority answers. Although the skewness and kurtosis of /u/ and /ɨ/ overlap each other, the mean skewness of the /u/ stimuli is 5.76 and the mean Kurtosis is /89.01/, and for /ɨ/ the majority stimuli skewness is /9.54/, whereas the Kurtosis is 207.3.

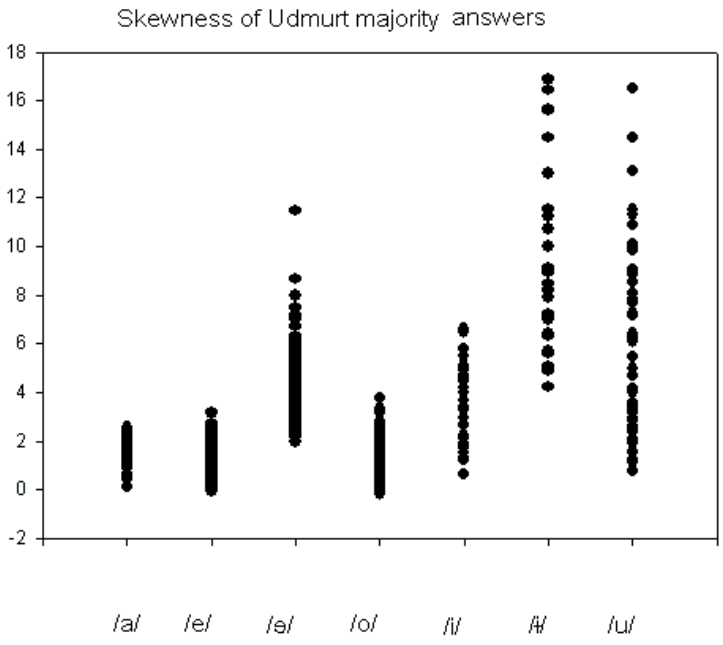


Figure 19.The categorisation of Turku Vowel Test stimuli, as plotted against skewness in Udmurt response by majority category

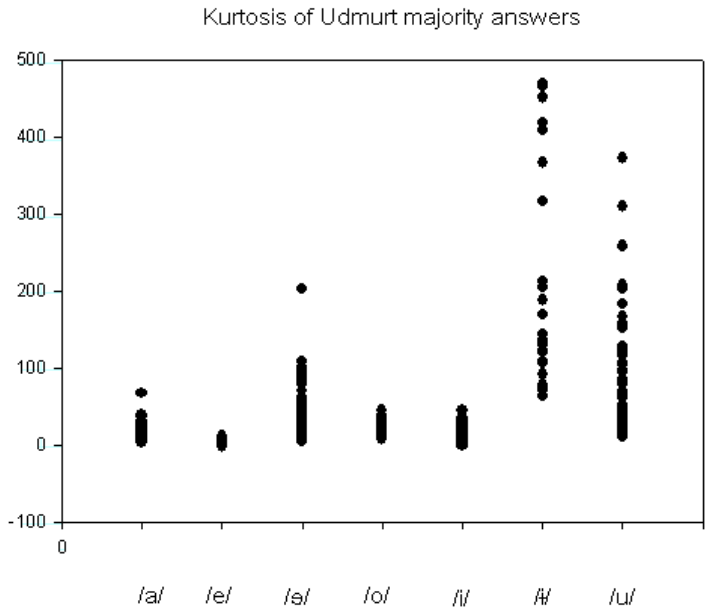


Figure 20. The categorisation of Turku Vowel Test stimuli as plotted against kurtosis in Udmurt response by majority vowel category

The relationship between individual vowel categories, acoustic attributes and their relationship to phonetic/phonological descriptions is not discussed in Study I. The relationship between vowels and their phonetic features is not studied systematically in this thesis because they are not crucial for demonstrating the general principles in the use of spectral moments and formants. Various phonetic features are compared, for example, by Lammela (2004).

3.3. Statistical evaluation of vowel identification

The role of acoustic attributes in categorisation of /i/ and /ɨ/ have been studied in study by Savela and Pikkanen (2004). In that study the identification of these vowel categories was studied by changing F2 and F3. Furthermore possible role of spectral moments was studied by using multiregression models. Vowel categories are assumed to be linear in F1 and F2 space. This means that two areas can be divided with straight lines to two areas using one or two formants, and the categories are linearly separable. Rosner lists a number of studies that show the linear boundaries between vowel space: German (Hose, Langner & Scheich, 1983; Traunmüller & Lacerda, 1987), Swedish (Carlson, Granström et al., 1970; Traunmüller & Lacerda, 1987), Dutch and Turkish (Traunmüller & Lacerda, 1987), Japanese (Akagi, 1993), and Russian (Karnickaya, Mushnikov, Slepokurova & Zhukov, 1975). The alternative parameters have also been proposed, as shown in Chapter 2.

Three analyses were conducted on the nature of vowel identification. The general effects of the acoustic attributes in Udmurt were examined in *Analysis 1*, while different sets of acoustic attributes are compared within four languages in *Analysis 2*. Finally, the effects of these attributes are compared in *Analysis* 3 in order to investigate their relative importance in identification.

3.3.1. Statistical analysis 1: The nature of identification and goodness rating in Udmurt back vowels

As shown above, the vowel charts of Udmurt and Komi (which belong to the same Permic, Finno-Ugric family) are divided into two areas on the grounds of F2. The central vowels indicated that the identification results cannot be explained by the formants. The hypothesis was that non-prototypical vowels are identified on the basis of spectral moments, whereas the prototypical stimuli are identified on the basis of formants (F1 and F2). Four statistical analyses were used to test the hypothesis. In the first two analyses the explaining variables were *formants* and *formants + spectral moments*, and the testable factors were all /u/ -responses. In the last two analyses the same explaining variables were compared with the goodness rating responses. The responses are described in the following set of figures (Figure 21, Figure 22).

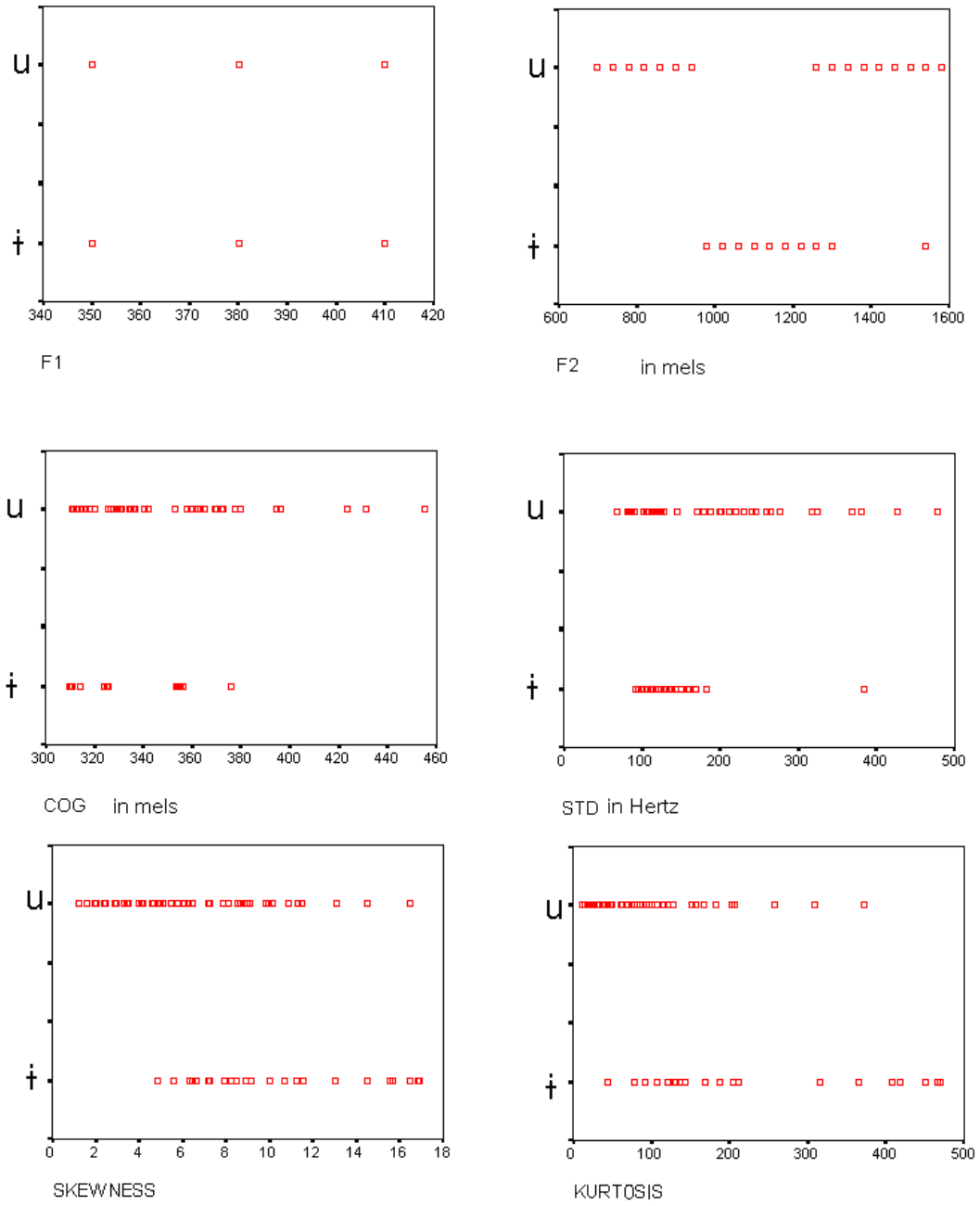


Figure 21. The /i̱/ and /u/ answers as functions of different spectral moments.The CoGs are described in mels, the STD in Herz and the skewness and kurtosis as coefficient values.

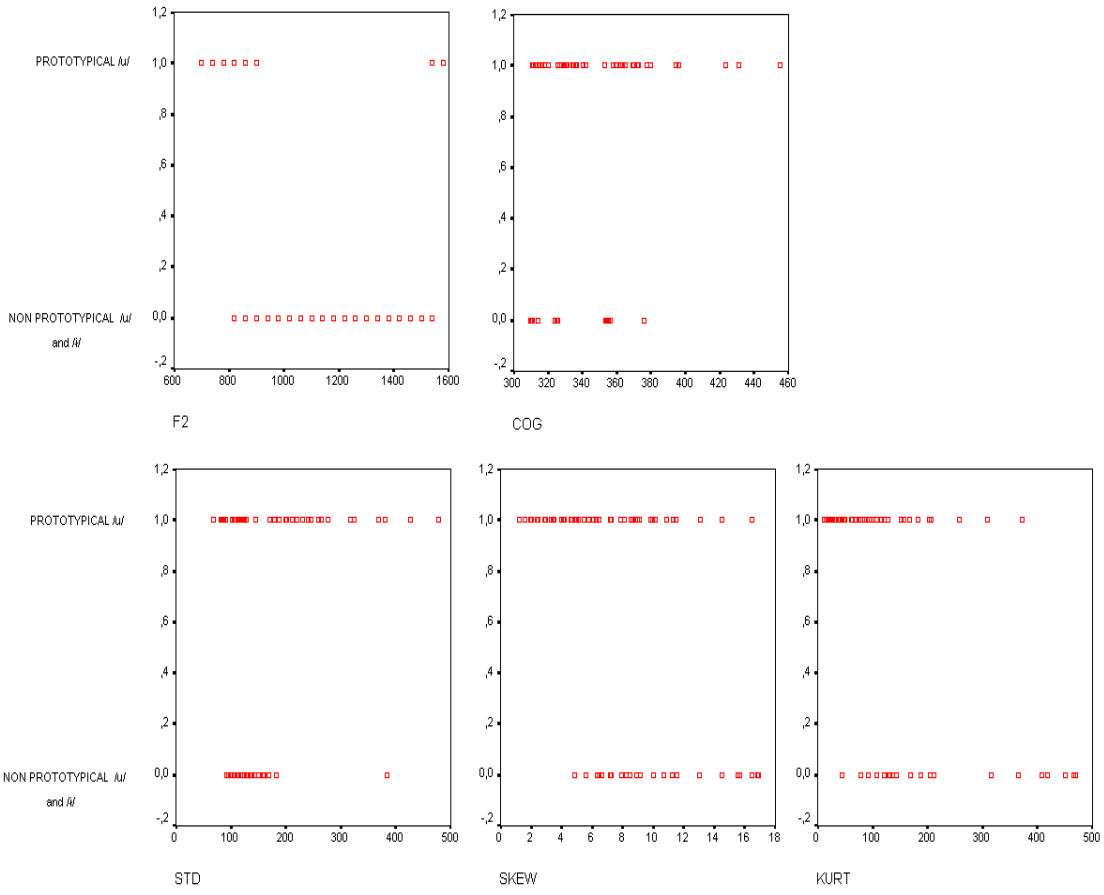


Figure 22. The prototypical /**u**/ and other answers as functions of different spectral moments
The CoGs are described in mels, the STD in Herz and the skewness and kurtosis as coefficient values.

3.3.1.1. Methods

Four statistical analyses based on the binary logistic regression model (using enter mode) were computed for stimuli categorised by a majority of listeners as /u/ or /ɨ/. The binary regression model is a statistical tool used to examine the relationship between observed responses and the various explanatory factors. Multi-nominal regression analyses have been used in phonetic research in terms of pattern recognition frameworks (e.g. Maddox, Molis & Diehl, 2002b; Nearey & Kiefte, 2003; Lammela, 2004) .

The methodology used in the present study is a binominal logistic regression model that is used for data in which the measured variables are binary and any kind of independent variables. Logistic regression is used to predict a measured variable on the basis of continuous and categorical variables. In Analysis 1 of Study 1 the continuous variables are different spectral measures. The models give the percentage variance in the dependent variable (the modal answer of the particular stimulus) explained by the variable (an individual spectral measure). The impact of these predictor variables is explained in terms of odd ratios.

In the first analysis the stimuli that were categorised by the majority as /u/, were given the value of 1, and otherwise 0. The described data was named as *the identification of /u-ɨ/ data*. In the next two analyses the selection criteria was the goodness rating of /u/ -category. If the stimuli were considered as /u/ and were rated above 4, they were given a value of 1, otherwise 0. The described data was named *the goodness rating of /u/ data*. Two models were used for the spectral attributes in the analyses: 1) the F1F2 -model in which the two lowest formants were used, and 2) the F1F2 + SM –model in which the spectral moments were also used.

3.3.1.2. Results and discussion

The results of fitting the binary logistic regression models are shown in Table 3. The binary logistic regression demonstrates whether the same phenomenon is described by: 1) the distribution of /u/ -responses and 2) the explaining variables of the model. If the explaining variables fit the observed distribution of responses significantly better than the null model, the model becomes significant. The tests used to show the efficiency of the models included the Chi-square model, which indicates whether the model is better than the null model (a model without the added variables). The goodness of fit elaborates the appropriateness of the models and, finally, Wald statistics explain the significance of the individual measured variables.

Table 3. Strength of models in Udmurt /ɨ - u/ identification
The first column shows the chi-square value, which indicates the general fit of the variables in the model (omnibus test on the variables within). The next column (DF) presents the degrees of freedom and the third column (sig.) presents the omnibus significance of the model compared to the null model.

	Identification of /ɨ - u/			Goodness rating /u/		
	Chi square	DF	sig.	Chi square	DF	sig.
The F1F2 model	0,077	2	.962	17,628	2	0,000
The F1F2 + SM model	36,128	6	.000	46,552	6	0,000

The results show that, in Udmurt, the identification of vowels /u/ and /ɨ/ was predicted not only on the basis of formants, whereas the goodness rating was predicted without adding the spectral moments to the model. The relative fits of spectral attributes are presented in Table 4. Although adding the spectral moments improved the model, the hypothesis that formants are enough to explain the goodness ratings, holds.

Table 4. The statistical significance of different spectral attributes in the identification of Udmurt vowels
The Wald Chi-square demonstrates the effect of the explaining variable in the overall fit of the model. The DF gives the degrees of freedom of the variable and the p gives the significance of the factor. The constant term is used in the evaluation of the model but it is not interested in the present study.

	VOWEL NAME /ɨ - u/		
	Wald Chi-square	DF	P
F1	,380	1	,537
F2	,571	1	,450
COG	0,039	1	,844
STD	,111	1	,739
SKEWNESS	2,379	1	,123
KURTOSIS	5,240	1	,022
CONSTANT	22,342	1	,138

The idea that the vowel perception of oral vowels can be based on other acoustic features found support in the figures and in the statistical analysis. Formants do not explain the results if only simple linear logistic methods are used. When modelling the data, the simplest model is considered to be the best, so the results show that in case of vowel stimuli, the formants do no explain the result in terms of linear boundaries. However, the other spectral attributes do exhibit this pattern. The high skewness and kurtosis values showed a relationship to the categorisation pattern /ɨ/, which means that they can act as decision criteria for identifying that category, whereas the formant-based model does not, as long as the vowel perception is understood to be based on pattern recognition theory at all.

The result may be related to the differences in vowel categorisation between individual subjects. The following figure (Figure 23) demonstrates the situation for different listeners.

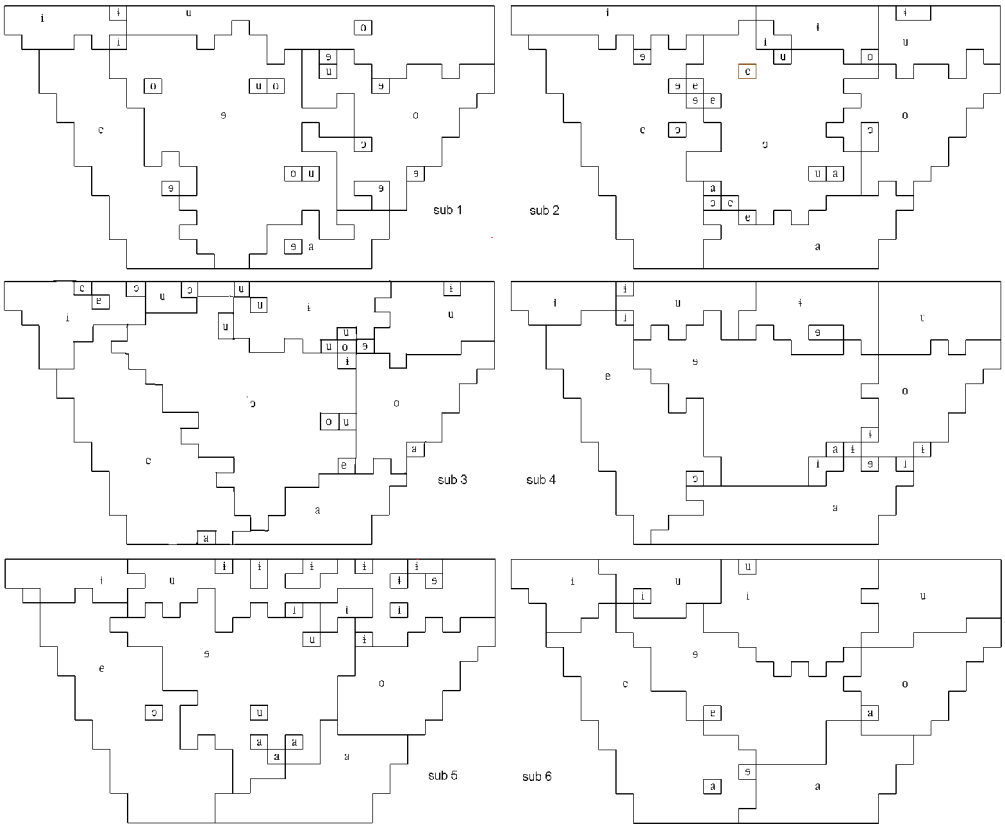


Figure 23. The vowel charts of six Udmurt subjects

The results showed that the double /u/ pattern can be found in three of the six subjects. The same phenomenon can also be seen in the Komi subjects in Janne Savela ś Master ś thesis (1999). In that study, eight out of 30 subjects exhibited such a phenomenon in their vowel space, and this phenomenon may be caused by several factors. The orthography of Udmurt and Komi reflects the phonological analysis of the languages fairly well (e.g. Rédei, 1978). Subjects were asked to identify whether the sound was <y> or another offered category. There was a possibility that the poor /u/ would be considered as /u/ in the palatalised environment (where the F2 of the vowel is usually raised) (Kuznetsov & Ott, 1987; Kuznetsov & Ott, 2001). This could explain why stimuli within the area of /y/ are identified as /u/, however, the pattern is different if the identification charts of Russian listeners are observed (Figure 24).

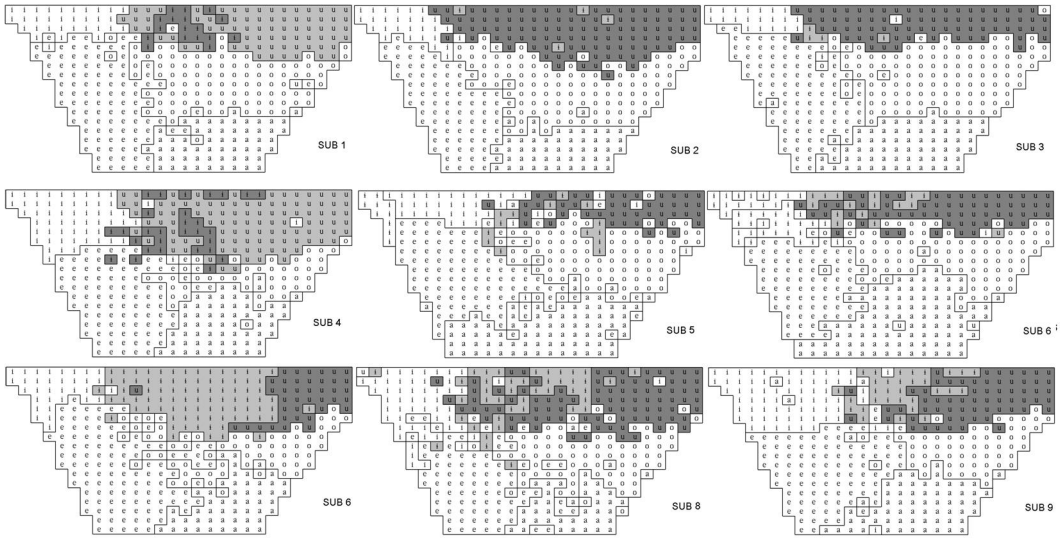


Figure 24. Results of nine Russian subjects in TVT database
Subjects were asked to choose between six vowel systems.

There were two orthographic symbols in Russian, <y> and <ы>, that are also used in Udmurt, although in Russian the <ы> is an velarised variant of the /i/ phoneme and the pattern is different than it is in Udmurt. Only subject 1 had a result that was similar to that of the Udmurts, which may be because the Russian < ы > is closer to the Polish /ɪ/, which is a retracted front vowel, not a central vowel. Its prototypical area is further back than the sound described by the same letter in Russian.

On the other hand, the more fronted Udmurt /u/ could reflect the Russian <ю>, which also has the same values of F2 as [y] sounds (Kuznetsov & Ott, 1987; Kuznetsov & Ott, 2001). Again, in terms of F2, the palatalised /u/ reaches the F2 of /y/, yet in Russian there is no interruption in the areas of < ю > and <y>, so from this perspective the pattern of Udmurt is different to Russian. There is no reported evidence on the effects of the palatalisation of sounds in Udmurt or Komi. Neither Iivonen (for Udmurt (Iivonen & Harnud, 2005)) or Savela (for Komi (Savela, 1999; Savela, 2000)) tested this difference.

Romanian contains /y/ in carefully uttered loanwords from French. For the Romanians this could lead to the interpretation of /u/ sounds with high F2 similar to [y], if they somehow manage them with their /u/. However, this could mean that the other /u/ is an allophone based on the articulation rather than on acoustic information like formant values. In terms of the acoustic properties of sounds, this pattern may be due to non-formant based similarity among the sounds in particular sound classes.

The vowel system with non-rounded central vowels is fairly common in world languages (10.7% based on a sample of a UPSID database of 451 languages)

(Maddieson & Disner, 1984). The problem is however, that /ɯ/,/ɪ/,/ɨ/ are not clearly differentiated in phonetic descriptions of languages. For example, Turkish and Japanese /ɯ/ are different (the Turkish vowel is actually closer to /ɨ/, whereas the Japanese /ɯ/ is more like /ʉ̟/ (Okada, 1999). This difference is, however, beyond the scope of this study.

3.3.2. Statistical analysis 2: Acoustic parameters in the identification responses of Finnish, German, Spanish and Czech

Finnish, German, Spanish and Czech were chosen for closer evaluation on the basis of the number of subjects and their similar vowel patterns. Spanish and Czech have typical five-vowel systems and Finnish and German have eight-vowel systems (four rounded and four non-rounded vowels).

3.3.2.1. Methods: Statistical procedure

In contrast to binominal regression models, there is more than one categorical variable in multi-nominal regression – one category serves as the contrast category. The results indicate whether the change in that particular variable is related to the change in the categorisation of the stimuli.

Multi-nominal logistic regression models are a statistical tool for analysing the effects of different attributes on phonetic performance (e.g. Maddox, Molis & Diehl, 2002; Nearey & Kiefte, 2003). The idea is that the vowel space can be divided into distinct areas that represent the different vowel categories. This idea is originally based on visual pattern recognition models and in that context it is called "general recognition theory". This means that each category is based on decision boundaries in different parameters (Ashby and Maddox 2005). The statistical likelihood to the particular category is maximal in the centre of the category and becomes smaller towards the boundary areas of the category. The symbolic process requires awareness (at least testing using forced-choice procedures). The idea is firstly to find the boundaries between categories. In contrast, the goodness ratings that are suggested to represent the associative (indexical) level of recognition are studied using goodness ratings, the idea of which is to show that the acoustic criteria for the vowel boundaries are different from the criteria for categorisation. Studies comparing many languages using this type of methodology have not been made and, furthermore, spectral moments have not been used in models.

For the four selected languages, identification frequencies were analysed by multimodal logistic regression models on the category's frequencies, with the goodness rating results used as weights. The analysis was conducted using the SAS programme in which the GENMOD was used. The first model (F1F2 - model) had two formants with linear and quadratic effects as the explaining parameters, and the primary response variable in a model was the identification frequency of each vowel. Explanatory variables included

standardised frequencies of formants and their quadratic effects. The range of formants in the standardisation was transformed into a 0 – 1 scale, considering the highest value on the axis as 1 and the lowest as 0, with formants in between receiving intermediate values. The interaction between F1 and F2 was included in the model, but all interactions with quadratic effects were excluded from analysis.

In the second model (F123 model), the explanatory variables were the three lowest formants. The explanatory variables in the third model (F12 + SM model) were the two lowest formants, their quadratic effects, their interaction, and the spectral moments (CoG, Std, normalised skewness and normalised kurtosis). In the fourth model (SM model) the spectral moments were used as the only explanatory variables.

3.3.2.2. Results and discussion

The four vowel representation models were compared using multimodal linear logistic regression in order to evaluate the categorisation data of different languages. The results for different stimuli sets with different spectral models (F1F2 model, F123 model, F12 + SM model, SM model) were computed, see (Table 5). The results were unequivocal. The SM model was the least accurate in all four languages and the F1F2 and F1F2F3 models fitted the data less accurately than the F1F2 + SM model. This supports the hypothesis on spectral moments as an additional acoustic attribute in the boundaries between vowel categories in perceptual vowel space.

Table 5. Strength of different models in terms of AIC (Akaike Information Criterion) in four languages
This table presents the fit of the logistic regression models. The AIC can be used to test the most suitable model for certain data. It is based on *the maximum log likelihood method* that is used for comparing models with different numbers of explaining variables. Lower AIC values indicate that the model is more suitable for the data.

The measured factors of the model	*F12 model*	*F1F2F3 model*	*F1F2 + SM model*	*SS model*
Czech	*7436*	*7380*	*7289*	*9083*
Spanish	*14400*	*14365*	*14237*	*19010*
Finnish	*58257*	*58021*	*57334*	*66589*
German	*16993*	*16787*	*16213*	*20465*

A closer look at the effects of different explanatory acoustic attributes was computed, see Table 6.

Table 6. The strength of different acoustic attributes in different languages
This table shows the explanatory values of the acoustic attributes in terms of *degrees of freedom* (DF) and *Wald chi-square* criteria, which explains the strength of a particular variable in the fit of the model. The stars indicate the significance of the variable in the model (p = * 0.01, ** 0, 001, *** 0, 0001).

	Finnish *Wald* *DF chi-square*	*German* *Wald* *DF chi-square*	*Spanish* *Wald* *DF chi-square*	*Czech* *Wald* *DF chi-square*
F1	*7 213.8127****	*7 216.8520****	*4 39.8363****	*4 116.4449****
F2	*7 141.8083****	*7 158.6876****	*4 31.1802****	*4 17.9350***
F1^2	*7 175.8987****	*7 258.2312****	*4 22.5571**	*4 113.8235****
F2^2	*7 107.5945****	*7 78.8158****	*4 10.92610**	*4 69.6901****
*F1*F2*	*7 418.3894****	*7 282.9083****	*4 79.2397****	*4 212.6890****
CoG_mel	*7 349.2433****	*7 206.8328****	*4 4.6664*	*4 18.6774***
Std_mel	*7 387.3565****	*7 211.1792****	*4 23.5295****	*4 51.4294****
Skewness	*7 107.1468****	*7 120.8113****	*4 32.8662****	*4 11.2006**
Kurtosis	*7 195.7297****	*7 131.0069****	*4 21.2497****	*4 13.3093**

The identification responses of the synthetic stimuli in Finnish are explained more by the CoG and Std than by the single formants F1 and F2 (although the interaction between the formants is the most effective variable), and German identification responses show a similar pattern. The quadratic effect of the formants was the most significant in the Spanish identification responses, whereas the CoGs had no effect. In the Czech identification responses, the vowel category boundaries reflected both the formants and the other acoustic attributes, but the other attributes were reflected to a lesser degree.

3.3.3. Statistical analysis 3: The prototypicality and acoustic properties of the stimulus

In order to compare the effects of prototypicality in identification, the same analysis was run for a prototypical set of stimuli. A stimulus was determined to be prototypical if its responses received more than 85% of the total answers to a certain stimulus.

3.3.3.1 Methods: statistical procedure

The identification responses for the four languages were divided into three subsets on the basis of the unanimity of identification; stimuli with >85% unanimity were treated as prototype areas (Figure 25). A test for stimuli with <85% unanimity was also computed but not reported, because the aim was to show that the prototypical stimuli are mostly independent from the categories. The dissociation between the non-prototypical and prototypical areas was shown in the case of Udmurt.

3.3.3.2. Results and discussion

Multimodal logistic regression models for prototypical subsets of data were computed, as shown in Table 7 and Table 8. The model explains how the observed categorisation frequencies (the distribution of responses into different vowel categories) can be predicted on the basis of a model ś chosen acoustic parameters. The models with spectral moments fitted the data better than those models without them (although all models were significantly fitted), and differences existed between languages with similar vowel systems.

In *Analysis 3* the languages were shown to differ in the way they used the spectral attributes in terms of identification and goodness rating. In some languages prototypes are identified mostly on the basis of the first two formants (F1 and F2). In Spanish the vowel prototypes are in peripheral areas avoiding central vowels.

Some languages, such as Finnish and German, prefer areas that are distinguished on the basis of formants and other acoustic attributes. In Finnish and German the CoG reflects vowel roundness, while in languages with unrounded back vowels, such as Udmurt, the primary criterion for identification is the combination of different spectral attributes. No acoustic attributes were observed in Czech that reflect the prototypicality of the stimulus.

Table 7. Fit of different spectral attributes (AIC - criterion) for models
The AIC can be used to test the most suitable model for particular data. A low AIC value indicates a more suitable model than a higher one. The results are related to the number of observations in different languages; therefore, absolute numbers in different languages should not be compared.

	Good (>85 %)	
	The F1F2 model	*The 12 + SM model*
Czech	*656.492*	*627.667*
Spanish	*1534.599*	*1487.184*
Finnish	*4646.093*	*4563.251*
German	*1510.341*	*1331.457*

Table 8. The strength of different acoustic attributes in the prototypical stimuli of different languages
This table shows the explanatory values of the acoustic attributes in terms of *degrees of freedom* (DF) and *Wald chi-square* criteria, explaining the strength of a particular variable in the fit of the model. Stars indicate the significance of the variable in the model (p = * 0.01, ** 0, 001, *** 0, 0001).

	Finnish	*German*	*Spanish*	*Czech*
	Wald	*Wald*	*Wald*	*Wald*
	DF chi-square	*DF chi-square*	*DF chi-square*	*DF chi-square*
F1	*7 38.4987****	*7 19.2973***	*4 58.7952****	*4 0.8299*
F2	*7 26.4464***	*7 17.1936**	*4 14.5807*	*4 3.8947*
F1^2	*7 25.1831***	*7 18.7764***	*4 57.3997****	*4 4.5521*
F2^2	*7 23.9151***	*7 18.2720***	*4 11.7342**	*4 3.5113*
*F1*F2*	*7 42.1667****	*7 22.4314***	*4 19.4973***	*4 8.6203*
CoG_Hz	*7 21.5337***	*7 39.9764****	*4 3.6752*	*4 6.7355*
Std_mel	*7 33.6387****	*7 29.6208****	*4 2.2449*	*4 8.7623*
Kurtosis	*7 26.8179***	*7 15.4585**	*4 43.7962****	*4 0.5831*
Skewness	*7 31.4532****	*7 16.9596**	*4 53.0836****	*4 1.7843*

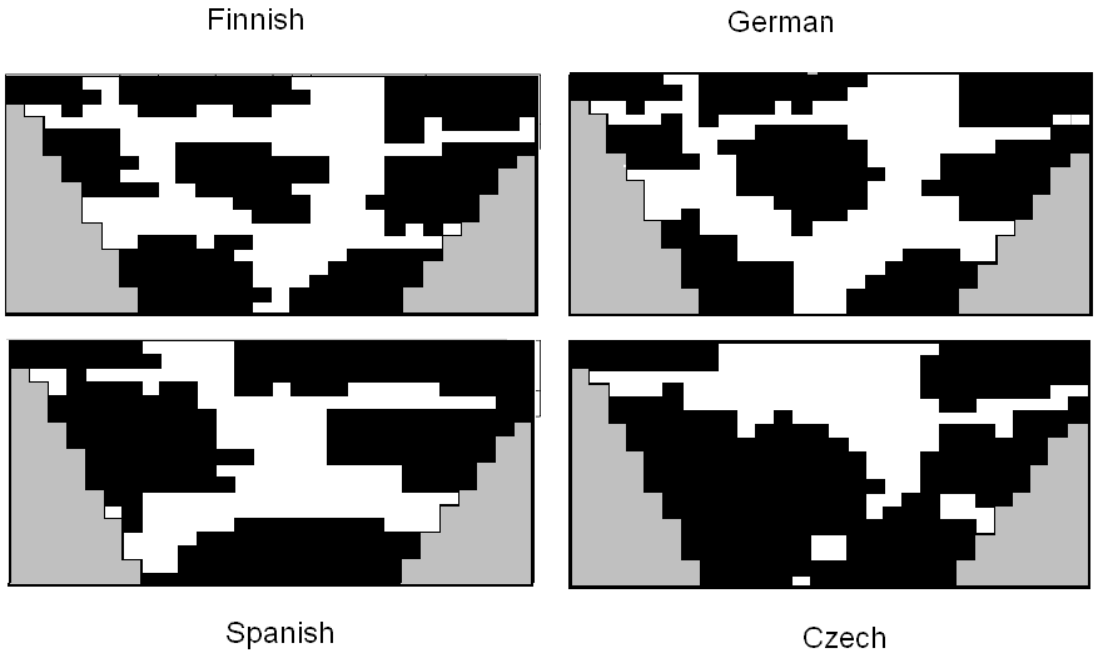


Figure 25. Selected stimuli in *Analysis 3*
The prototype areas (over 85% unambiguity) are presented in black.

3.4. CONCLUSIONS

Together, the vowel charts and six statistical models showed that vowel boundaries and prototypical stimuli in the identification and goodness rating tasks were based on different acoustic attributes in different languages (in terms of perceptual vowel space). The results of Udmurt showed that prototypes and identification criteria are based on the use of different spectral attributes (formant prototypes or spectral attributes). This indicates that in the

vowel stimuli with same formant values, the symbolic knowledge of the listener is reflected systematically in additional spectral attributes, not just in F1 and F2. However, the goodness of stimuli can be explained on the basis of formant-based prototypes in terms of linear combinations of F1 and F2. This result goes against the idea that vowel boundaries mirror the distribution of formant-based vowel prototypes, as would be suggested if formant-based prototype rules were correct. The possible relationship of the present results to the articulatory gestures is not used in the present study, because the relationship of the articulation and the acoustic properties cannot be modelled directly for present purposes. However, since the relationship between sounds and the muscular activity behind sound can be interpreted to have many to one relationships, the formant values themselves can be interpreted as crucial feedback information for the speaker/listener while delivering speech signals (Guenther et al. 1998). On the basis of different vowel charts and the results of statistical analyses, it is difficult to speculate about the mechanism behind the vowel identification. This means that there should be a larger variety of possible measures in modelling the vowel identification of different languages. In some languages, most prominently in Spanish, the vowel prototypes seem to be distributed evenly in the formant space and the vowel boundaries are also dependent upon them in terms of the closest vowel prototype. In some systems the vowels reflect additional acoustic attributes, such as CoG in German and Finnish and Std in Czech. It appears that these languages do not use the formant-based prototype matching procedure to identify the stimuli.

In terms of the general architecture of the vowel system, several theories have predicted the location of prototypes in efficient vowel spaces. According to the model of Lijencranz and Lindblom (1972), the distance between speech sounds is based on the acoustic dispersion of vowel prototypes, that is, the maximal acoustic distance between them in terms of the Euclidean database. In TVT data, only Spanish reflected this type of pattern in the identification responses. On the other hand, according to the *focalisation-dispersion theory*, the vowels are placed within some focal areas. In those areas some of the formants are situated close to each other (Schwartz, Boe, Vallee & Abry, 1997b).

The final question, however, is whether the acoustic attributes belong to the language experience of the listener. It must be noted that the values of the acoustic attributes may be related to the particular set of vowels in the TVT data set. The stimuli of TVT are fairly untilted, with large amplitude on the higher areas of the vowel spectrum. In the TVT data set the CoGs may be relatively high, compared to more tilted vowel sets. Unfortunately, it was not possible to compare these properties between different vowel sets.

Having said that, it can be argued that these kind of acoustic attributes are also part of a speaker ś identity, since they are a result of both the speaker ś source (i.e. the glottis) and their filter (i.e. the vocal tract) (Milenkovic & Forrest, 1988). However, that is not a problem with the present type of vowel identification experiment. The role of the spectral moments (the non-formant

measure) in vowel identification has been investigated by Sakayori et al. (Sakayori, Kitama, Chimoto, Qin & Sato, 2002), who argue that the slope of the amplitude between the formants gives directly speaker-related information on the vowel categorisation. Study I presented a more detailed picture of the use acoustic attributes, and showed that different acoustic attributes (i.e. formants and spectral moments) can be used in the categorisation in various ways in different vowel classes.

4. STUDY II: THE EFFECT OF STIMULUS ASSIGNMENT

It has been shown in the field of experimental psychology that the discrimination of characteristic features in an acoustic stimulus depends not only on the acoustic distance between two stimuli but also on the order in which they are presented (Repp, Healy & Crowder, 1979; Cowan & Morse, 1986; Repp & Crowder, 1990).

Cowan and Morse (1986) found evidence for the general psychophysical phenomenon of *neutralisation* in vowel discrimination. The preceding stimulus was neutralised in the direction of the schwa-vowel (mid-central vowel). This means that the difference between the sounds of an AX pair (A=standard and X=deviant) was perceived to be larger if the A was more central than the X, and smaller if the X was more central than the A. On the contrary, in the study led by Repp (1990) no general tendency towards schwa was found. The results showed that the dependence of contrast effects on the stimulus order (i.e. neutralisation) was based on the inner structure of each vowel category.

The effects of prototypicality in pre-attentive sound discrimination have been shown to reflect the acoustic distance between two vowels (Näätänen, 1997). If the standard is prototypical and the deviant is non-prototypical, the elicited MMN response is less than if both the standard and the deviant are prototypical. Study II examines the effects of stimulus order and prototypicality on the attentive and pre-attentive level of sound discrimination. A similar experiment was conducted by Ikeda et al. (Ikeda, Hayashi, Hashimoto, Otomo & Kanno, 2002) in which the prototypicality of the Japanese [e] was reflected in the MMN. A prototypical standard stimulus elicited a larger discrimination response in MMN than a non-prototypical one. Study II examines this phenomenon using three different vowel pairs that belong to three different vowel categories (/e/-like, /ø/-like, and /o/-like stimuli). The prototypical stimuli were determined on the basis of the Komi and Finnish vowel identification tests.

4.1. Experiment 1: Attentive discrimination

4.1.1. Methods

Subjects

Ten students from the University of Turku (mean age 22.3 years, age range 20-30, six females) participated in the experiment. All were native speakers of Finnish without any command of Komi and none reported any hearing problems.

Stimuli and apparatus

Six steady-state stimuli 8 (Table 9) were synthesised with the HLSyn (high-level parameter speech synthesis system, version 1.0 Sensimetrics). The formant values were chosen on the basis of Savela ś previous data (Savela, 2000), which represented the prototypical vowels of Komi (/ɛ/, /ə/, /o/) and Finnish (/e/, /ø/, /o/). If more than one stimulus had been ranked as prototypical, their mean F1 and F2 were used. The duration of the vowels was 350 ms and the fundamental frequency (F0) of the stimuli was 100 Hz. The onset and offset of the stimuli were smoothed by a ramp of 10 ms.

Table 9. Formant and spectral moment values of the stimuli used in Study II (in Hz)

	Vowel type					
	Finnish			Komi		
	/e/	/ø/	/o/	/ɛ/	/ə/	/o/
F1	430	430	460	550	505	565
F2	2400	1775	800	2350	1350	825
F3	3200	2800	3200	3200	2800	3200
CoG	385	336	417	540	423	551
Std	447	196	182	655	239	190
skewness	5, 8	5, 6	0, 4	3, 7	3, 6	-0, 4

The vowel pairs, each consisting of one Finnish and a corresponding Komi vowel, formed three groups: front /e - ɛ/, central /ø – ə/, and back /o – o/. Thus, the acoustic distance was largest for the /ø - ə/ pair, and about equal for the other two pairs. However, the stimuli in the /e- ɛ/ pair were more different phonetically than those in the other two pairs because the Komi /ɛ/ is not a clear allophone of /e/ in Finnish. The acoustic/auditory distance is presented in the following table (Table 10).

Table 10. Difference between stimuli in different acoustic/auditory scales

	/e- ɛ/	/ø - ə/	/o - o/
Formant distance	*103*	*207*	*98*
CoG distance (mel scale)	*150*	*90*	*126*
Difference in Std (mel scale)	*89*	*21*	*10*
Difference in skewness	*2, 1*	*2, 0*	*0, 8*

A total of six blocks were used for each experiment. In three of the blocks the Komi vowel served as the standard and the Finnish vowel as the deviant, and vice-versa in the other three. There were 51 deviants and 350 standards in

each block, meaning that the probability of the deviant was 0.125. The stimuli were randomly presented in each block with a silent inter-stimulus interval (ISI) of 400 ms and the blocks were presented in a different order for each listener.

A reaction time session that lasted about 45 minutes took place after the mismatch negativity (MMN) recordings (see Experiment 2). Subjects were asked to press the response button with their right index finger as soon as possible after hearing a deviant stimulus in the stream of standard stimuli. Reaction time was measured from the onset of the stimulus and responses that were shorter than 100 ms and longer than 2500 ms were immediately excluded. In addition, all reaction times that were longer or shorter than the subject's mean reaction time by three times the standard deviation (1.1% of all reaction time measures) were treated as outliers and were excluded from further analysis.

4.1.2. Results and discussion

The overall error rate was 1.2%. Misses accounted for 2.1% and the false alarm rate was 0.4%, with no statistically significant differences in these rates between the conditions. The individual mean reaction times (Table 11) to correct responses were subjected to repeated measures analysis of variance (ANOVA).

Table 11. Reaction times for different vowel pairs in the discrimination experiment

	Standard Vowel type					
	Komi (non-prototypical standard)			Finnish (prototypical standard)		
	/ɛ/	/ɘ/	/o/	/e/	/ø/	/o/
Mean reaction time (Std.) (Errors %)	369 (44) (3, 3)	339 (58) (5, 1)	339 (46) (3, 9)	340 (55) (3, 1)	331 (42) (6,2)	348 (46) (3, 5)

Within-subject factors were the language of the standard vowel (Komi versus Finnish) and the vowel category (/e- ɛ/, /ø - ɘ/, /o -o/). The results indicated that the vowel category had a significant effect on the reaction time measure ($F(2, 18) = 4.403$ $p = 0.028$), which reflected the differences in the acoustic distance between the vowels of the different pairs. Post-hoc analyses (t-tests) showed that the reaction time measure was significantly shorter for the /ø - ɘ/ pair than for the /e- ɛ/ ($p = 0.038$) and /o -o/ ($p = 0.019$) pairs. Although there was no significant main effect for the language of the standard vowel, this factor interacted with the vowel type [$F(2, 18) = 4.959$ $p = 0.019$], which was due to the /e- ɛ/ pair. When the Finnish /e/ served as the standard, the

reaction time was shorter than when the Komi /ɛ/ was the standard (t (9) =2.840, p= 0.019).

The results showed, firstly, that the acoustic distance between the deviant and standard stimulus has an effect on reaction times. However, assigning the stimuli as standard or deviant affected the response time when one vowel of the pair was in the border area between two vowel categories (/e- ɛ/ pair). Based on these results, it is suggested that it is easier for the listener to perceive the difference between two vowels when the deviant is unfamiliar than when it is not.

4.2. Experiment 2: pre-attentive discrimination (event-related potentials)

Vowel discrimination was studied by using ERPs. Subjects heard the same stimuli used in *Experiment 1*, but were instructed to ignore the vowels they heard instead of discriminating them. Specifically, the MMN was used to determine the pre-attentive discrimination of the members of each vowel pair.

4.2.1. Methods

Subjects and stimuli

The subjects and stimuli were the same as in *Experiment 1,* and the same block design was used. However, in order for the MMN response to obtain a sufficient signal-to-noise ratio, the number of stimuli per block was increased. Accordingly, there were 151 deviant stimuli and 900 standard stimuli in each block; the occurrence probability of the deviant stimulus was 0.144. The stimuli were presented in a pseudo-randomised order in the odd-ball paradigm (in which deviant stimulus is presented randomly among standard ones) using the Neurostim programme. During the session subjects watched a silent movie and were under instructions to ignore the vowel sounds, which were binaurally presented through earphones at a comfortable sound-pressure level (about 70 dB). The session took about 1.5 hours.

Equipment and procedure

The continuous EEG was recorded using a Braintronics 32-channel EEG amplifier connected to a Neuroscan EEG data acquisition and analysis computer. The signal was amplified and stored on the hard disk of a personal computer and Ag/AgCl cup electrodes were used. Three electrodes were placed in standard locations (Fz, Cz, and Pz) according to the 10-20-system. Six lateral electrodes (three on the left and three on the right) were positioned at non-standard locations placed equidistantly on the coronal line connecting the mastoids through Fz. These locations were labelled as L1, R1, L2, R2, LM (left mastoid), and Rm (right mastoid). Eye movements were recorded with two

electrodes, one placed at the outer canthus of the eye and the other at FPZ, while the reference electrode was placed on the tip of the nose. The electrode impedance was kept below 5 kΩ and automatic artefact rejection (+/- 70 µv) was applied. The amplifier bandwidth was set at 0.5 – 70 Hz, and a sampling frequency of 200 Hz was used. For ERP averaging, for the standard and deviant stimuli separately, a 450 ms epoch was used (including a 50 ms pre-stimulus baseline interval). The ERP epochs were digitally filtered off-line by a 1-30 Hz band pass filter. The ERPs were re-referenced to the average of the left and right mastoids. For each stimulus block, the ERP response to the standard stimulus preceding the deviant was subtracted from that of the deviant. The peak amplitudes of the resulting MMN were measured separately for each subject from the difference waveforms. The MMN mean amplitude for each subject was measured using a 30 ms window cantered on the grand-average of the mismatch negativity peak-amplitude latency (Figure 26) (Table 12).

Table 12. The MMN (mismatch negativity) amplitudes (in **µ**V) for each block

	Standard vowel type					
	Komi (non-proto typical)			Finnish (prototypical)		
	/e/	/ɘ/	/o/	/e/	/ø/	/o/
Mean amplitude MMN (std. Dev)	-1, 47 (1, 5)	-4, 09 (2, 4)	-2, 0 (1,6)	-2,62 (2,2)	-3,36 (2,1)	-2, 11 (1, 6)

No main effect was found for the language of the standard vowel. However, the vowel type did have a significant effect on the MMN amplitude [F (2, 18) = 9,332 p= 0.002], which was significantly larger for the /ø - ɘ/ than for the /e-ɛ/ (p= 0.012) and /o –o/ (p< 0.001) contrasts. There was no significant interaction between the language of the standard stimulus and the vowel type.

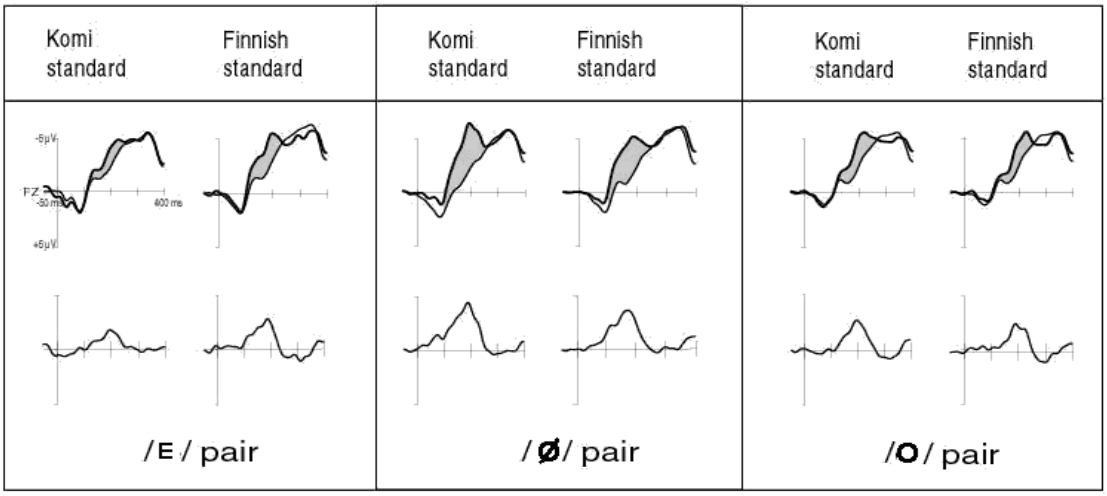


Figure 26. ERP waveforms for each block in the mismatch negativity experiment
The upper panel shows the waveform for the standard and deviant stimulus. The lower panel shows the subtraction waveform for the same block. The /e- ɛ/ blocks are abbreviated as /E/ pair, the /ø - ɵ/ blocks are abbreviated as /Ø/ pair and the /o –o/ are abbreviated as /O/ pair.

4.2.2. Results and discussion

The results showed no categorisation effects in the MMN amplitudes, but did show some effect of prototypicality in the reaction time. The reason for the difference between the two discrimination tasks could be the fact that the representations behind the MMN amplitudes are sensitive to the formants but not to the prototypicality of the stimulus. However, in the study by Ikeda et al (Ikeda, Hayashi et al., 2002) the MMN reflected the prototypicality of the sounds. The difference between the result of Study II and Ikeda ṡ study may be due to the shorter ISI (inter-stimulus interval) that Ikeda used, which may increase the category-based discrimination and decrease the acoustic continuity in the discrimination set.

On the other hand, our behavioural study (*Experiment 1*) showed a difference between the continua. This may be due to the difference in standard deviation (Std) between the stimuli, which was noticeably larger in the /e- ɛ/ block than in the other continua. This result is in line with the earlier experiment, which showed higher sensitivity to the Std for Finnish subjects than for other tested acoustic attributes (Study I). However, this effect seems to be post-perceptual or at least its pre-attentive effects are less sensitive due to inter-subjective variations typical for vowel categories (Aaltonen et al. 1997).

5. STUDY III: FORMANTS AND SPECTRAL SHAPE IN THE DISCRIMINATION OF VOWELS

Discrimination of the vowel continua can be affected by several factors that define the result of the discrimination experiment (in AX paradigm), the two most important of which are the inter-stimulus interval (ISI) and the preceding stimuli (Macmillan, 1987). Variation in any of these factors can affect the difference in detection of the stimuli. The internal structure of the category also affects the vowel discrimination (Iverson & Kuhl, 2000). Study III explores three other questions: the role of language experience, the linguistic description of the sound and the spectral manipulation of acoustic attributes. This is explored in the discrimination of two vowel continua: /æ – e/ and /æ – ø/. These continua have the same distance in the Euclidean vowel space but different distance in all of the spectral moments: the centre of gravity (CoG).

The MMN studies for vowels have demonstrated the effects of the pitch and the amplitude enhancement in vowel perception (Jacobsen, Schroger & Alter, 2004). In Jacobsen's study the general amplitude level of the stimulus did not affect the processing of vowels on the level of the MMN, yet these stimuli did not vary in spectral moments. In these terms the variation in the acoustic attributes determining identification (Study I) is not tested directly. In Study III the end point vowels on the two continua belong to different categories in Finnish, but to same categories in Spanish (see Study I). It can be demonstrated, therefore, that in the attentive discrimination of vowels the spectral moments are not dependent on the subject's native language.

The vowel stimuli in *Experiment 1* are chosen on the basis of Finnish identification responses. In *Experiment 2* the role of two acoustic attributes, the formants and the centres of gravity (CoG and other moments), are studied in Finnish and Spanish. In *Experiment 3*, the vowels are studied by using masked stimuli (white noise added to the stimulus) in order to alternate the most prominent spectral moments in the stimulus. *Experiment 4* compares the mismatch negativity (MMN) recordings in the pre-attentive discrimination of the vowels.

The aim of Study III is to show that any acoustic attribute available can be used in the attentive vowel discrimination, whereas the pre-attentive discrimination mostly reflects Euclidean formant distance.

5.1. Experiment 1: Stimulus selection

In order to select the stimuli for the discrimination task and for the MMN recordings in *Experiment 1*, 10 native Finnish speakers were asked to identify the stimuli in two vowel continua, /æ – e/ and /æ – ø/. Both continua consisted of 11 vowel stimuli in steps of 15 mels and were synthesised with a Klatt synthesiser (Klatt, 1980). In the /æ – e/ continuum F1 varied from 484 to 655 Hz, and F2 varied from 1756 to 1933 Hz. In the /æ – ø/ continuum F1 varied between 484 and 655 Hz and the F2 from 1592 to 1756 Hz. F3 was fixed at 2474 Hz in both continua and the duration of the stimuli was 380 ms. The fundamental frequency (F0) rose firstly from 100 to 120 Hz until 125 ms and then declined to 80 Hz during the rest of the stimulus.

Eight stimuli were chosen for the discrimination test on the basis of the identification test (Figure 27). The boundary stimuli were selected on the basis of the PROBIT regression analysis (Finney, 1971). In the PROBIT analysis the regression curve is fitted to the observed response. The point on the scale that receives a probability of 0.50 is considered the category boundary. The stimulus 6.4 (on the 1–11 scale, in which the numbers corresponded to the number of stimulus on continuum meaning 15 mel between two stimuli) in the /æ – ø/ continuum was a boundary stimulus and the slope of the curve was –2.254. Furthermore, the stimulus 6.6 on the /æ – e/ continuum was a boundary stimulus and the slope of the curve was -3.04.

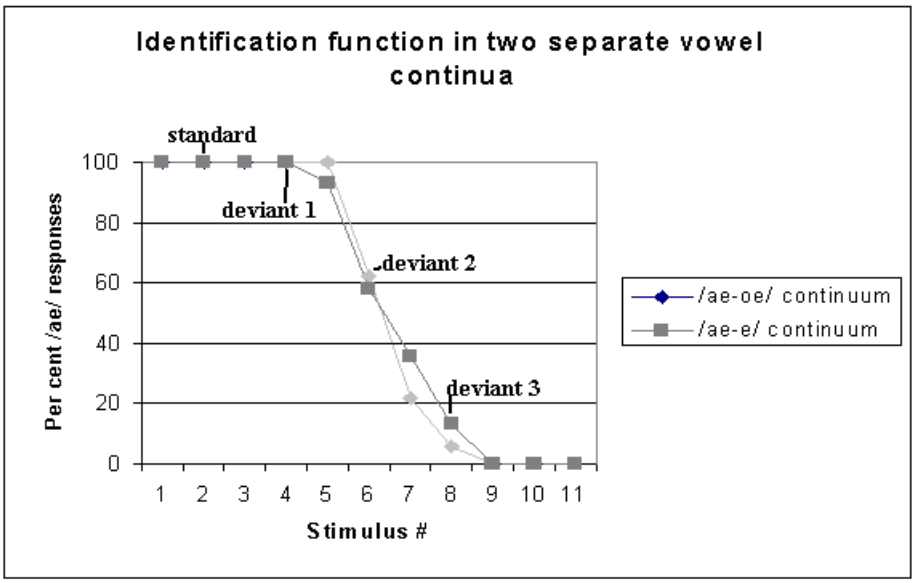


Figure 27. Identification functions for the vowel continua in Finnish

The first deviant was a within-category stimulus, the second one was located at the category boundary and the third deviant belonged to an adjacent category. Since the category boundary slopes in both continua were identical, the effects of the boundary (for example, differences in sensitivity) were expected to be similar.

5.2. Experiment 2: Attentive discrimination

5.2.1. Methods

Subjects

Fifteen students from the University of Turku participated in the experiment (mean age 23.6 years, range 20–25 years, all females). They were all native Finnish speakers with no reported hearing abnormalities and two of them were left-handed. Seven Spanish exchange students also participated (mean age 23.5 years, range 20–26 years, one left-handed, two females). All were native Spanish speakers and had not learned Finnish (the mean length of their stay was 2.2 months). One Finnish subject was excluded from the analysis due to a lack of MMN data.

Stimuli and procedure

The stimuli were chosen on the basis of the identification test (Table 13).

Table 13. Formants (Hz) and the centre of the gravity (mel) of the stimuli, the standard deviation of stimuli (Hz) and coefficients of skewness and Kurtosis Experiment 2

	/æ –e/ continuum				/æ –ø/ continuum			
Stimulus	Standard	Deviant 1	Devian 2	Deviant 3	Standard	Deviant 1	Deviant 2	Deviant 3
F1	636	601	566	532	636	601	566	532
F2	1773	1808	1843	1879	1739	1706	1672	1640
F3	2474	2474	2474	2474	2474	2474	2474	2474
CoG	861	823	792	765	843	776	715	662
Std	616	617	613	624	587	531	474	435
skewness	1.74	1.86	2.02	2.07	1.86	2.22	2.62	2.95
kurtosis	3.11	3.40	3.89	3.89	3.85	6.00	8.86	11.64

The standard stimulus /æ/ and 3 deviant stimuli, their acoustic distance from the standard being small (33 mels), medium (66 mels), or large (100 mels), were presented in the odd-ball paradigm. The differences between the standard and deviants were similar in both continua but varied in terms of the CoGs (center of gravity). The CoGs were measured using a frequency domain from 0 Hz to 5500 Hz for a stimulus. The differences in the CoGs were considerably larger in the /æ – e/ continuum than in the /æ – ø/ continuum (Table 14). There were two pseudo-random blocks, each of which consisted of 275 (82%) standard stimuli and 20 (6%) presentations of each of the three deviants. The inter-stimulus interval (ISI) was 400 ms. Subjects were asked to push a button when they heard a sound deviating from the stream of the standard stimuli. The session took 20 minutes.

Table 14. The Euclidean formant distance and the CoG difference (in mels), Experiment 2

	/æ – e/			/æ –ø/		
Deviant	Small	Medium	Large	Small	Medium	Large
Distance in formants (in mels)	32	64	97	35	67	104
Distance in CoGs (in mels)	55	111	149	103	185	249
Distance in Stds (in mels)	-1	7	-3	56	113	152
Difference in skewness	-0,12	-0,26	-0,33	-0,36	-0,76	-1,06
Difference in kurtosis	-0,29	-0,78	-0,78	-2,15	-4,81	-7,79

Table 15. Mean reaction times (in ms) and miss rate (in percent) for different distances between deviants and standards (in mels), Experiment 2

Finnish subjects						
	/æ – e/ continuum			/æ – ø/ continuum		
Distance	Small	Medium	Large	Small	Medium	Large
MeanRT (s.d.)	528 (63)	456 (63)	439 (65)	483 (65)	408 (53)	417 (54)
Miss rate	13, 2	1, 1	0	9, 3	0, 7	0
Spanish subjects						
Distance	Small	Medium	Large	Small	Medium	Large
MeanRT (s.d.)	625 (102)	534 (70)	517 (60)	582 (80)	493 (57)	478 (59)
Miss rate	49	10	1	36	2	1

5.2.2. Results and discussion

The results for the different stimuli are presented in Table 15.

Finnish subjects

The overall error rate for the Finnish-speaking subjects was 2.5%, the false alarm rate was 0.96% and the miss rate was 4.0%. Reaction times longer or shorter than three Std from the mean (a total of 0.4% of the responses) were excluded from the analysis. The individual miss rates were subjected to a repeated measures analysis of variance (ANOVA). The factors were the continuum type (/æ –e/, /æ – ø/) and the acoustic distance between the standard and deviant stimuli (short, medium, long). The acoustic distance had a significant effect on the miss rate [$F(2, 26) = 19.277$ $p<0.01$], with a larger acoustic distance decreasing the miss rate (Table 15).

The individual mean reaction times for each stimulus (Table 15) were subjected to ANOVA, using the same model as for the miss rates. The reaction times significantly reflected the acoustic distance between the stimuli [$F(2, 26) = 55.292$, $p < 0.001$]. The shortest acoustic distances (short) were discriminated significantly more slowly (505 ms) than the larger ones, medium: 432 ms ($p < 0.001$); large: 428 ms ($p < 0.001$). The type of continuum had a significant effect on reaction times [$F(1, 13) = 21.441$, $p < 0.001$]: a change in the /æ – ø/ continuum was discriminated significantly faster (474 ms) than in the /æ – e/ continuum (436 ms).

Spanish subjects

The overall error rate of Spanish-speaking subjects was 7.5% and the FA rate was 1.0%. The miss rate for the deviants was 16.5%. One subject was excluded from the analysis due to insufficient data. Reaction times longer or shorter than three Std from the mean (2.0% of responses) were excluded from the analysis. Individual miss rates were subjected to ANOVA, with the factors being the vowel continuum type (/æ –e/, /æ – ø/) and the acoustic distance between deviant and standard (short, medium, long). The acoustic distance had a significant effect on the miss rate [$F(2, 12) = 32.450$, $p < 0.001$], and a longer acoustic distance decreased the miss rate (Table 15). The continuum type had a significant effect on the miss rate [$F(1, 6) = 6.503$, $p = 0.043$]: the /æ – ø/ contrasts were perceived with significantly fewer misses (13.0%) than the /æ – e/ contrasts (20.0%).

The individual mean reaction times for each stimulus (Table 15) were subjected to ANOVA. The acoustic distance between the standard and deviant was reflected in the reaction times [$F(2, 12) = 37.437$, $p < 0.001$]. The Spanish-speaking subjects discriminated the shortest acoustic distance

(small) significantly more slowly (mean = 604 ms) than the larger acoustic distances (medium: mean = 513 ms; long: mean = 497 ms). The effect of the vowel continuum was also significant [F (1, 6) = 8.617, p = 0.026]. The changes in the /æ – e/ continuum were perceived more slowly (558 ms) than those in the /æ –ø/ continuum (518 ms).

Inter-group comparison

ANOVA was used to determine the difference between the groups using the subjects ' native language as a between-subjects factor. There was a significant difference in the miss rate between the subject groups [F (1, 19) = 22.156, p < 0.001]. The miss rate of the Spanish-speaking subjects was 16.5%, compared to 4% for the Finnish-speaking subjects. The language background had a significant interaction with the acoustic distance [F (2, 38) = 23.036, p < 0.001]. Finally, the Spanish subjects made more errors in the /æ – e/ continuum than in the /æ – ø/ continuum, resulting in significant interaction between the subject group and the continuum type [F (1, 19) = 5, 421, p = 0.031]. The /æ – e/ continuum was perceived significantly less accurately than the /æ – ø/ continuum. There was a significant difference between the reaction times in the two subject groups [F (1, 19) = 9.688, p = 0.006]. The Spanish-speaking subjects were slower (538 ms) than the Finnish ones (455 ms), and the acoustic distance had a significant effect on the reaction times [F (2, 38) = 90.833, p < 0.001]. The targets with the smallest distance (554 ms) from the standard were detected significantly slower than the targets with medium (473 ms, p < 0.001) and large distances (463 ms, p < 0.001). However, the language background had no interaction with the stimulus type [F (2,19) = 0,034, p=0.857] or the acoustic distance between standard and deviant [F (2, 38) = 1, 929 p < 0.159].

The continuum type had an effect on the reaction times for both subject groups. Acoustic distance affected reaction times similarly, being independent from the subjects 'language background. Overall, the Spanish-speaking subjects were slower than the Finnish-speakers and made more misses. This could be due the fact that Finnish subjects are more experienced with stimuli in that area in vowel space. The discrimination of the two continua is based on the CoGs (or other moments) in the present study. It is not possible, theoretically, to say which spectral moment would be the best candidate for the salient feature. However, it can be seen that the difference between /e/ and /ø/ in Finnish vowel charts seems to be CoG and it is probable, therefore, that the Finnish listeners use it as the

boundary criterion. In the case of Spanish, the perceptual boundaries seem to reflect the distance of formant prototypes, so the vowel discrimination may reflect any of the spectral moments. The use of spectral moments as an auxiliary cue to the vowel identification boundaries may contribute to the general speed of the various listener groups.

5.3. Experiment 3: Discrimination in noise

In *Experiment 2* the attentive discrimination of vowels seemed to be affected by the CoGs (and other moments) instead of formants. In order to validate this observation, the CoGs were manipulated by adding white noise into the stimuli. The difference in discrimination speed between the two vowel continua was expected to disappear, as the differences in the CoGs were less salient when the stimuli were embedded in noise.

5.3.1. Methods

Subjects

Twelve students from the University of Turku participated in the experiment (mean age 24.3 years, range 20 – 40 years, six females). All were native Finnish speakers, with no reported hearing abnormalities and one subject was left-handed.

Stimuli and procedure

The stimuli used in *Experiment 1* were embedded in white noise with the same RMS amplitude as the stimuli (+ 0 snr) (Figure 28). The differences in CoG did not exist or were barely detectable in both continua if Flanagan's difference limens for CoGs were correct (Flanagan, 1955) (Table 16). Two blocks of stimuli were presented in the same odd-ball paradigm as in *Experiment 1*, with each block consisting of 206 (82%) standards and 15 (6%) presentations of each of the deviants. The subjects were asked to push a button when they heard a sound that deviated from the stream of the standard stimuli. The session took 20 minutes.

Table 16. The distances (standard - deviant) between the CoGs (in mels), Stds (in Herzs) and skewness and kurtosis coefficients in different signal-to-noise conditions

	/æ –e/			/æ –ø/		
Deviant	Small	Medium	Large	Small	Mediu m	Large
Quiet						
CoG	55	111	149	103	185	249
Std	-1	7	-3	56	113	152
skewness	-0,12	-0,26	-0,33	-0,36	-0,76	-1,06
kurtosis	-0,29	-0,78	-0,78	-2,15	-4,81	-7,79
0 dB						
CoG	0	21	33	14	32	70
Std	-16	-20	-31	-22	-3	-3
skewness	0,00	0,00	-0,01	0,00	-0,11	-0,11
kurtosis	-1,33	0,01	0,01	0,000	0,013	0,013

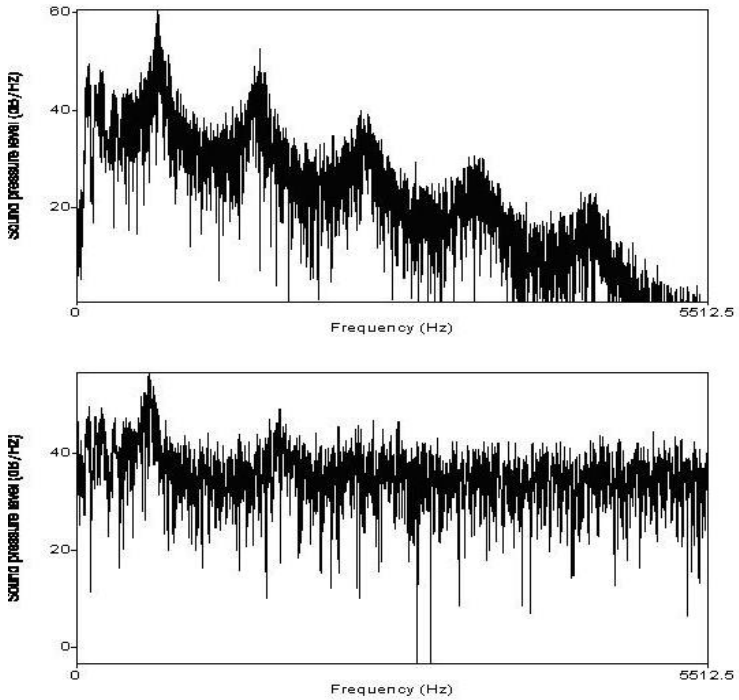


Figure 28. Same stimulus in quiet conditions and with noise
Upper panel: Stimulus in a quiet situation
Lower panel: -0 dB white noise added to the stimulus

5.3.2. Results and discussion

The overall error rate for the Finnish-speaking subjects was 8.0% and the false alarm rate was 1.23%, while the miss rate for the deviants was 16.8%. The individual miss rates were subjected to ANOVA. The factors were the continuum type (/æ –e/, /æ – ø/) and the acoustic distance between standard and deviant (small, medium, large). Only the acoustic distance had a significant effect on the miss rate (Table 18) [$F(2, 22) = 77.333$ $p<0.01$] and this was decreased by larger acoustic distances.

The individual mean reaction times for each stimulus (Table 17) were subjected to ANOVA by using the same model as for the miss rates. The reaction times significantly reflected the acoustic distance between the stimuli [$F(2, 22) = 25.333$, $p < 0.001$], with the smallest acoustic distances (small) being perceived significantly more slowly (580 ms) than the larger ones [medium: 516 ms ($p < 0.001$); large: 498 ms ($p < 0.001$)]. The continuum type had no effect on subjects 'reaction times.

Table 17. Reaction times (in ms) and miss rates (in percent) with signal-to-noise ratio being 0 dB, Experiment 3

	/æ –e/ continuum			/æ –ø/ continuum		
Deviant	Small	Medium	Large	Small	Medium	Large
Quiet	579 (62)	520 (61)	499 (58)	580 (89)	512 (97)	496 (101)
0 dB	42, 8 (21,9)	6, 7 (6, 9)	0, 6 (1, 9)	44, 4 (26, 1)	5, 6 (17, 6)	1, 1 (3, 8)

As expected, the difference between the continuum types disappeared when there was a small difference between CoGs of standard and deviant (only 10 mels), although the formant difference remained larger (33 mels). The conclusion drawn from this is that attentive perception may be affected by the whole spectral shape in quiet conditions but by the formants when noise is added into the stimuli. Therefore, those stimulus properties that provide the most distinctive information about the stimuli are used in discrimination. The differences in standard deviation did not explain the result if noise was added to stimulus.

5.4. Experiment 4: ERP recordings

In this experiment, ERP recordings were used in order to compare attentive and pre-attentive discrimination of the same stimuli. If the pre-attentive discrimination used the same properties of the stimuli as the attentive discrimination in quiet conditions, no difference should exist between continua in the MMN amplitudes. If the CoGs (and other spectral moments) are the basis for the discrimination, the deviant stimuli on the /æ – ø/ continuum should elicit larger MMN amplitudes than those in the /æ –e/ continuum.

5.4.1. Methods

Subjects

The same Finnish subjects were used as in *Experiment 1*.

Stimuli and procedure

The stimuli used in the reaction time measurements without noise were also used for the MMN recordings. In order to increase the signal-to-noise ratio of the recordings, 120 of each deviant stimulus and 1440 standard stimuli were used in each block. The stimuli were presented in a pseudorandom order in the odd-ball paradigm using the Neurostim programme. The probability of occurrence of a deviant stimulus was 0.20. Subjects watched a silent movie during the session and were instructed to ignore the vowel sounds, which were binaurally presented through earphones at a comfortable sound-pressure level (about 70 dB). The session took about an hour.

The recording procedure of MMN was largely similar to Study II. The continuous EEG was recorded by using a Braintronics 32-channel EEG amplifier connected to a Neuroscan EEG data acquisition unit and an analysis computer. The signal was amplified and stored on the hard disk of a personal computer. Ag/AgCl cup electrodes were used, three of which were placed on the standard locations (Fz, Cz, and Pz) according to the 10-20-system. Six lateral electrodes (three on the left and three on the right) were positioned at non-standard locations placed equidistantly on the coronal line connecting the mastoids through Fz. These locations were labelled as L1, R1, L2, R2, LM (left mastoid), and Rm (right mastoid). Eye movements were recorded with two electrodes, one placed at the outer

canthus of the right eye and the other at Fpz, while the reference electrode was placed on the tip of the nose. The electrode impedance was kept below 5 kΩ and automatic artefact rejection (+/- 70 µv) was applied. The amplifier bandwidth was set at 0.5–70 Hz, and a sampling frequency of 200 Hz separately for the standard and deviant stimuli. A 450 ms epoch was used for ERP averaging (including a 50 ms pre-stimulus baseline interval). The ERP epochs were digitally filtered off-line by a 1.6–30 Hz band pass filter and the ERP responses to the standard stimulus preceding each deviant were subtracted from that to the deviant for each stimulus block. In contrast to Study II the MMN amplitudes and latencies were measured from the difference waveforms (from the time window 150–210ms) separately for each subject.

5.4.2. Results and discussion

The individual MMN amplitudes and their latencies for each condition were tested with a one sample t-test. This test showed that all of the amplitudes differed significantly from zero (in all condition $p<0.01$, Boniferroni corrected for multiple comparisons) (Table 18) (Figure 29), and they were subjected to ANOVA. The factors were the continuum type (/æ – e/, /æ – ø/) and the acoustic distance between the deviant and standard stimuli (small, medium, large). Neither the acoustic distance [$F(2, 26) = 1.975$, $p=0.173$] nor the continuum type ($F(2,26)=2.075$, $p= 0,193$] had any effect on the MMN amplitude. The latency of MMN amplitude was significantly affected by the acoustic distance between the standard and deviant stimuli [$F(2, 26) = 5.598$, $p = 0.010$], but not by the continuum type [$F(2,26) = 2.075$, $p=0.173$]. The deviant with the smallest acoustic distance to the standard stimulus had a significantly longer mean latency (227 ms) compared to the larger ones (medium: mean = 200 ms, $p = 0.023$; large: mean = 199 ms, $p = 0.009$).

Table 18. MMN amplitudes (**µ**V) and latencies (ms), Experiment 4
All amplitudes differed significantly from 0 **µ**V ($p < 0.001$).

Continua	/æ – e/			/æ –ø/		
Deviant	Small	Medium	Large	Small	Medium	Large
MMN latency (ms)	220	202	197	233	199	206
Std. deviation	(37)	(49)	(35)	(26)	(42)	(37)

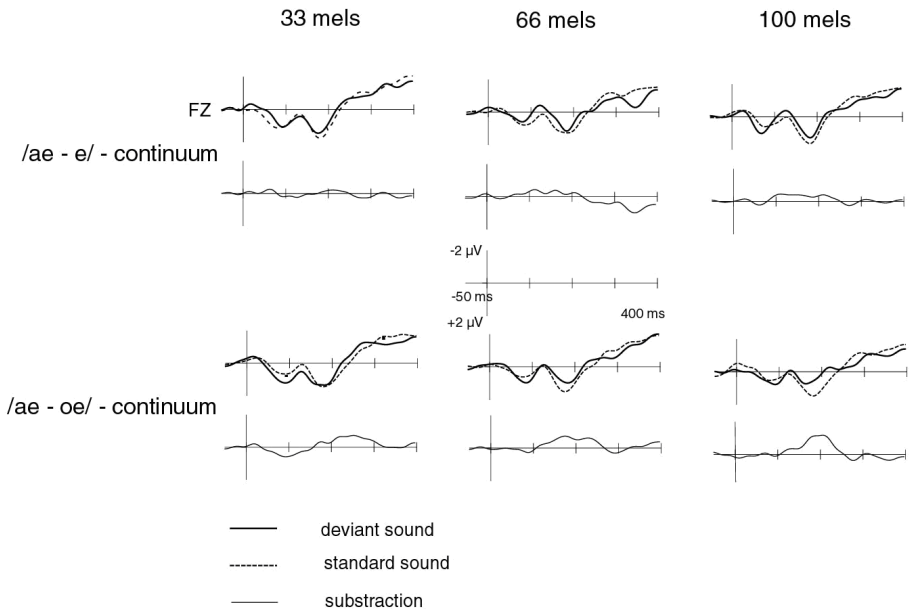


Figure 29. ERP waveforms for different continua
For each continuum the waveform for standard and deviant stimuli are shown in the upper panel and the subtraction waveform in the lower panel.

5.5. Discussion

The vowel discrimination yielded three results. The auditory representation based on the attentive discrimination was shown to be independent of the linguistic background (whether the subjects were Finnish or Spanish). The Spanish subjects were generally slower in RTs than the Finnish subjects.

Both subject groups seemed to pay attention to the most distinctive acoustic attribute. In other words, their criteria in attentive discrimination tasks were based on the maximal dissimilarity. This was indicated by their reaction times, which reflected the differences between centres of gravity. However, the Finnish subjects changed the discrimination criteria in the noise condition. On the other hand, the MMN responses reflected the formants and the results showed that the MMN latencies reflect the formants, not the spectral moments.

This can be considered to indicate the process in which auditory representations in long-term memory are activated, and this means that they are not based on afferent features of the stimulus. The MMN is known to require a representation to the standard stimulus, which can be shown by different features of MMN that cannot be interpreted without the concept memory-based trace (Näätänen and Winkler 1999). The results indicate that the preference for formants can be found in the early stage of vowel processing. This results from the fact that the use of spectral moments as additional acoustic attributes in discrimination were independent from their use as criteria in identification (see Study I). Study III is in line with studies by Jacobsen (2002), which showed that the MMN reflects the formant differences more accurately than the amplitude level differences of the same stimuli. The reason why the difference was significant at the level of MMN latencies but not in the level of MMN amplitudes will not been discussed in more detail in here.

6. GENERAL DISCUSSION

The theme of this thesis was to present evidence to support the idea that vowel categorisation follows the semiotic principles in which the emergence of indexes (i.e. the experience based inner structures of vowel categories) governs the use of acoustic attributes in sound perception.

The aim was to show that vowel identification is primarily indexical, not symbolic (that is, it is not based only on the holistic phoneme system). This was done by comparing the role of acoustic attributes at the level of identification of different phonetic categories. The use of two types of acoustic attributes in attentive and pre-attentive vowel perception was investigated using synthetic vowel stimuli. The identification responses (Study I) and the discrimination responses (Studies II and III) were compared on the basis of formants and spectral moments.

The statistical analysis of the TVT data (Study I) showed that the spectral moments were used as an additional attribute, with the formants being the main criteria in vowel identification. However, in the Udmurt data the spectral moments were necessary to explain the identification response, since the role of formants was not significant in the statistical analysis (Study I: Analysis 1). The use of acoustic attributes as identification and goodness rating criteria varied in different languages (Study I: Analysis 2 and Study I: Analysis 3). Thus the acoustical structure of areas for particular vowel categories in different languages must be different, respectively.

In Study II the mismatch negativity (MMN) results indicated that the pre-attentive processing of vowels was based mainly on the formants. Instead, the attentive discrimination was sensitive to the prototypicality differences based on the stimulus order. This may possibly have resulted from the relative short ISI (inter-stimulus interval) and the long duration of the stimuli (see e.g. Ikeda, Hayashi et al., 2002). These parameters can be examined more thoroughly in the future.

Study III showed that attention modulated the use of spectral moments in the discrimination response. The attentive discrimination by both Finnish and Spanish subjects reflected the spectral moment measure (CoG),

although the Finnish subjects were more sensitive to smaller differences than their Spanish counterparts. The results suggested that the acoustic attributes used in the identification task did not thoroughly explain the discrimination responses. The manipulation of the CoG by adding noise into the stimulus was reflected in the discrimination responses. The automatic (i.e. pre-attentive) discrimination mainly used the formants in Finnish subjects, and the other spectral moments also supported the result.

The statistical analysis of the spectral moments showed them to be an additional acoustic attribute used in the vowel identification. Nevertheless, in some sets of identification response they reflected the identification performance efficiently. In addition, spectral moments may be used in particular situations as emergent criteria to identify the stimuli, however, this type of dissociation between formants and spectral moments is important in order to understand how the vowels are perceived. It is concluded that neither the formants nor the spectral moments can be the only explaining factor in perceiving vowels.

There are two possible solutions to explain the use of acoustic attributes in vowel identification. The first solution is based on the exemplar-based models, in which the acoustical information of vowels can be stored in the episodic memory of vowel traces. This exemplar-based solution has been proposed by Shestakova et al. (2002). The other possible solution is based on the distinctive features that possibly direct the vowel identification to certain areas of similarity (Stevens, 1980). According to Stevens, the speech perception is quantal, i.e. in some areas a small articulatory difference is reflected by a large acoustic difference and vice versa.

The pure Euclidean space has not been unanimously accepted to explain the location of prototypes in the vowel space (Polka & Bohn, 2003). The peripheries are preferred for the prototypes in some studies but not in others. This thesis suggests that the use of spectral attributes may reflect the development of vowel perception in early childhood, while on the level of adult listeners, the iconic acoustic attributes may be used to group the vowels to phonetic classes. Thus the phonetic classes as themselves may not serve as symbolic criteria that structure the vowel space, as has traditionally been suggested. For example, for Finns the CoG may reflect the roundness whereas for the Udmurt it is the skewness. However, it cannot be said whether these attributes are also used actively at the level of words since only isolated steady-stated vowels have been investigated in

this thesis. Furthermore, the spectral moments may have their own role as an indicator of stress feature in vowels (Sluijter & van Heuven, 1996).

This thesis also suggests that the spectral moments may also lead to the emergence of new vowel categories by themselves, with the formants having no essential role in the process. On the basis of the identification results, the non-formant-based vowel targets within the vowel space were located, and those targets may serve as acoustic targets in vowel perception, indicating subregions within the vowel space.

Finally, some future prospects are presented on the basis of the results:
The first prospects are more theoretical:

1. This thesis favoured the idea of phonetic semiosis, in which the new categories can be found - not only on the basis of formants - but also on the basis of additional acoustic attributes. In attentive perception the spectral moments may serve as an only criterion of the phonetic quality (Ito, Tsuchida et al., 2001), although it is difficult to understand how this would happen on the pre-attentive level of vowel perception. Furthermore, the role of attention on the acoustic attributes behind other speech sounds (i.e. fricatives) cannot be explained solely on the basis of this thesis, which has concentrated only on vowels. This would be a fruitful research subject in the future. In general, the pattern recognition models provide a method to study the end result of the identification process and are therefore useful in comparisons of different languages (e.g. Nearey 1996).

2. It is possible to classify different combinations of the acoustic attributes. The differences in the use of acoustic attributes are essential, above all, in the identification of non-prototypical vowels. Besides the traditional formant-based classification, a new, more profound classification of vowel identification criteria would be useful. The preliminary classification of vowel systems can be seen already on the basis of this thesis, but because the results did not show a distinct pattern, the understanding should be developed further.

3. In language learning, the role of additional acoustic attributes (i.e. spectral moments) must not be ignored. For example, in developing applications for training perception and articulation of foreign speech sounds, it must be taken into consideration that vowel identification is not based only on the first two formants but on the combination of different acoustic attributes. In practice it could be studied whether the education

technology for speech sounds should include information about spectral moments if the differences with vowels systems are tested. The pattern recognition model used in present study – the difference between symbolic and indexical knowledge on speech sounds – can be used in the modelling of vowel identification. It should be considered whether the phenomenon that is planned to be taught to the listener is affected by symbolic or indexical knowledge of sounds, and what would be an appropriate model for each particular situation.

4. Vowels have different distribution of spectral moments in different languages. Compared to formants, the role of spectral moments seems to be rather suggestive. However, since their role is related, above all, to individual variation, the relationship between individual articulation habits and spectral moments should be investigated more profoundly. This could lead to a deeper understanding of the differences between languages.

5. The identification responses of TVT can be utilised when investigating sound change from historical, individual and social perspectives. This could be used in modelling the role of the listener in dialectology and in studying historical sound on the basis of perception. It provides a deeper perspective to understanding the role of the listener in the emergence of phonetic categories that may be affected by indexical and symbolic strategies of the subject. The logistic regression models can be used in this task by adding knowledge on these variants to the models. This aspect can lead to many practical applications in which the linguistic background of the listener can be analysed. The present study provided results for several languages, but more sophisticated dimensions of linguistic identities can be examined if more data is gathered.

The following perspectives concern the applications based on this thesis.

6. The location of prototypes (in terms of formants) and the use of additional attributes (spectral moments) may be used in the identification of listener-related differences between vowel spaces. Both of these features may reflect the individual strategies and personal experience with particular sound categories, which means that it may be possible to compare the identification performance by one subject with the model that should represent his or her individual personal experience. This could make it possible to compare the subject's identity in terms of the individual aspects in vowel recognition. Subjective differences may be greater if non-familiar vowels are used, because this evaluates the subject's pattern recognition

strategies. There are two situations that are crucial for the subjective differences. Firstly, the relationship between prototypical and non-prototypical tokens could change between speakers, and secondly, the use of particular spectral features may differ from one listener to another. Because vowel identification is a combination of familiarity and symbolic system, the individual strategy in building new categories could also be tested. This is most profound in a situation in which there is no consensus on the symbolic level of the particular language, as is the case with Russian.

7. This study may be used for evaluating pathological applications of phonetic systems. It can be used in order to guide therapy in which the linguistic abilities of the speaker are not normal. For example, it is adequate to provide information for subjects that are not able to articulate vowels normally (such as different modifications of the vocal tract) (e.g. Niemi, Laaksonen, Vähätalo, Tuomainen, Aaltonen & Happonen, 2002; Niemi, Laaksonen, Aaltonen & Happonen, 2004). Understanding the difference between symbolic and indexical levels of identification may help to evaluate patients ' phonetic and linguistic abilities. For example, it should be considered whether a change in formants is the only criterium for acoustic analysis of vowels (Rodman, McAllister, Bitzer, Cepeda & Abbitt, 2002; Erikson, Cepeda, Rodman, McAllister, Bitzer & Arroway, 2004). They found the spectral moments (especially standard deviation) to be stable cues for speaker identity. Different spectral moment patterns can also be used to synthesise the different speaker types. The use of spectral moments as the index of a speaker ś identity is a promising methodology in speaker-related studies and it could be also used in studies of speaker recognition by the listener. It could be used as an additional source of information in the synthesis, if seeking the differences between the speakers.

8. The results can be used to help understand the noise resistance of the phonetic signal. The results showed that the formant distance was used if the subjects were asked to discriminate the stimuli they perceived. This phenomenon can help to understand the core elements of the speech identification processes and may be used in packing of relevant information about the signal. Requests for this information have been crucial in examining the phonetic mode in the signal (Remez, Rubin et al., 1994). The idea is that speech can be seen as a specified pattern in study, that is, it is modal and based on its own mechanisms. Speech recognition technology should naturally imitate these elements in the speech stream for speech perception modelling.

In general, the understanding of the physical properties behind the phonetic concepts makes it possible to understand the differences between individual human interpretations of phonetic acts.

Bibliography

Aaltonen, O., Eerola, O., Hellstrom, A., Uusipaikka, E. and Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data, Journal of the Acoustical Society of America 101 (2): 1090-1105.

Aaltonen, O., Eerola, O., Lang, H. A., Uusipaikka, E. and Tuomainen, J. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity, Journal of the Acoustical Society of America 96 (3): 1489-1493.

Aaltonen, O., Hellstroem, A., Peltola, M., Savela, J. and Tamminen, H. (2008). Brain responses reveal hardwired detection of native-language rule violations, Neuroscience Letters 444: 56-59.

Aaltonen, O., Niemi, P., Nyrke, T. and Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum, Biological psychology 24 (3): 197-207.

Akagi, M. (1993). Modelling of contextual effects based on spectral peak interaction, Journal of the Acoustical Socety of America 93: 1076-1086.

Alivuotila, L., Hakokari, J., Savela, J., Happonen, R.-P. and Aaltonen, O. (2007). Perception and imitation of Finnish open vowels among children, naïve adults and trained phoneticians. 16th International Congress of Phonetic Sciences, Saarbrücken, Universität des Saarlandes.

Alivuotila, L., Savela, J. and Aaltonen, O. (2008). Kielitaustan vaikutus vokaaleja matkittaessa, Puhe ja Kieli/Tal och Språk/Speech and Language 28 (3): 129-140.

Ashby, F. G. and Maddox, W. T. (2005). Human category learning, Annual Review of Psychology 56 (1): 149-178.

Assmann, P. F., Nearey, T. M. and Hogan, J. T. (1982). Vowel identification: orthographic, perceptual, and acoustic aspects, Journal of the Acoustical Society of America71 (4): 975-989.

Assmann, P. F. and Summerfield, Q. (1989). Modelling the perception of concurrent vowels: Vowels with the same fundamental frequency, Journal of the Acoustical Society of America85 (1) 680 -697.

Aulanko, R., Hari, R., Lounasmaa, O. V., Näätänen, R. and Sams, M. (1993). Phonetic invariances in the human auditory cortex, NeuroReport 4: 1356-1358.

Bell, A. M. (1867). Visible Speech. London, Simpkin Marshall & Co.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B. and Sawusch, J. R. (1979). Some relationships between speech production and perception, Phonetica 36 (6): 373-383.

Bernstein, J. (1981). Formant-based representation of auditory similarity among vowel-like sounds, Journal of the Acoustical Society of America 69 (4): 1132 - 1144.
Bladon, R. A. and Lindblom, B. (1981). Modelling the judgment of vowel quality differences, Journal of the Acoustical Society of America69 (5) 1414-1422.
Boersma, P. and Weenink, D. (2001). Praat: doing phonetics by computer.
Butcher, A. (1974). 'Brightness,' 'Darkness' and the Dimensionality of Vowel Perception, Journal of Phonetics 2: 153 –160.
Carlson, R. and Granström, B. (1979). Model predictions of vowel dissimilarity, TMH-QPSR 3-4: 84-104.
Carlson, R., Granström, B. and Fant, G. (1970). Speech perception: some studies concerning perception of isolated vowels, STL-QPSR 2-3: 19-35.
Chiba, T. and Kajiyama, M. (1958). The vowel its nature and structure, Phonetic society of Japan.
Chistovich, I. A. and Chernova, E. I. (1986). Identification of one- and two-formant steady-state vowels: a model and experiments, Speech Comm 5: 3-6.
Cooper, F. S., Liberman, A. M. and Borst, J. M. (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech, Proceedings of the National Academy of Sciences Washington 37: 318-325.
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 (2): 500-507.
Cross, D. V., Lane, H. L. and Sheppard, W. C. (1965). Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception, J Exp Psychol 70: 63-74.
Deacon, T. W. (2003). Universal grammar and semiotic constraints. Language evolution. M. H. Christiansen and S. Kirby, London Oxford University Press: 111-139.
Diehl, R. L., Lotto, A. J. and Holt, L. (2004). Speech perception, Annual Review of Psychology 55: 149-179.
Diesch, E. and Luce, T. (2000). Topographic and temporal indices of vowel spectral envelope extraction in the human auditory cortex, Journal of Cognitive Neuroscience 12 (5): 878-893.
Eerola, O., Laaksonen, J.-P., Savela, J. and Aaltonen, O. (2003a). Perception and production of the short and long Finnish vowels: Individuals seem to have diffrent perceptual and articulatory templates. 15th Congress of Phonetic Sciences 3.-9. August 2003, Barcelona, Universitat Autònoma de Barcelona: 989-992.

Eerola, O., Laaksonen, J.-P., Savela, J. and Aaltonen, O. (2003b). Suomen [y /i] ja [y: /i:]-vokaalien tuotto havainto kokeiden valossa. Fonetiikan päivät, Akustikan ja äänenkäsittelytekniikan laboratorio, TKK, Otaniemi: 109-113.
Eerola, O., Savela, J., Laaksonen, J.-P. and Aaltonen, O. (2003). Keston vaikutus suomen [y] / [i] jatkumon kategorisointiin ja [i] vokaalien prototyypin havaitsemiseen. Fonetiikan päivät, Akustiikan ja äänenkäsittelytekniikan laboratorio, TKK, Otaniemi: 115-122.
Eggermont, J. J. (2001). Between sound and perception: reviewing the search for a neural code, Hearing Research 157 (1-2): 1-42.
Erikson, E. J., Cepeda, L., Rodman, R., McAllister, D., Bitzer, D. and Arroway, P. (2004). Cross-language speaker identification using spectral moments. FONETIK 2004 - The XVII Swedish Phonetic Conference, Stockholm, Department of Linguistics, Stockholm University.
Fant, G. (1960). Accoustic theory of speech production : with calculations based on X-ray studies of Russian articulations. 's-Gravenhage, Mouton & Co.
Fant, G. (1973). Speech Sounds and Features.
Finney, D. J. (1971). Probit analysis. Cambridge, Cambridge University Press.
Flanagan, J. L. (1955). Difference limen for vowel formant frequency, Journal of the Acoustical Society of America 27 (3): 613-617.
Flipsen, P. J., Shriberg, L., Weismer, G., Karlsson, H. and McSweeney, J. (1999). Acoustic characteristics of /s/ in adolescents, Journal of Speech, Language and Hearing research 42 (3): 663-677.
Forrest, K., Weismer, G., Milenkovic, P. and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data, Journal of the Acoustical Society of America84 (1): 115- 123.
Fowler, C. A. (1996). Listeners do hear sounds, not tongues, Journal of the Acoustical Society of America 99 (3): 1730-1741.
Gay, T., Lindblom, B. and Lubker, J. (1981). Production of bite-block vowels: acoustic equivalence by selective compensation, Journal of the Acoustical Society of America69 (3): 802-810.
Guenther, F. H., Hampson, M. and Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements, Psychological Review 105 (4): 611-633.
Harnad, S. (1987). Category induction and representation. Categorical perception: The groundwork of cognition. S. Harnad. New York, NY, Cambridge University Press: 535-565.
Hawks, J. W. (1994). Difference limens for formant patterns of vowel sounds, J.Acoust.Soc.Am 95 (2): 1074-1084.

Hellwag, C. (1781). De formatione loquale. Tübingen.
Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis of speech, Journal of the Acoustical Society of America 87 (4): 1738-1752.
Hose, B., Langner, G. and Scheich, H. (1983). Linear phoneme boundaries for German synthetic two-formant vowels, Hearing Research 9 (1): 13-25.
Iivonen, A. and Harnud, H. (2005). Acoustical comparison of the monophtong systems of Finnish, Mongolian and Udmurt, Journal of International Phonetic Association 35 (1): 59-71.
Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K. and Kanno, A. (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference, Neuroscience Letters 321 (3): 133-136.
IPA (1949). Principles of the International Phonetic Association.
Ito, M., Tsuchida, J. and Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception, The Journal of the Acoustical Society of America 110 (2): 1141-1149.
Iverson, P. and Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism?, Perception and Psychophysics 62 (4): 874-886.
Jacobsen, T., Schroger, E. and Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli, Psychophysiology 41 (4): 654-659.
Jacobsen, T., Schroger, E., Horenkamp, T. and Winkler, I. (2003). Mismatch negativity to pitch change: varied stimulus proportions in controlling effects of neural refractoriness on human auditory event-related brain potentials, Neuroscience Letters 344 (2): 79-82.
Jakobson, R., Fant, G. and Halle, M. (1961). Preliminaries to speech analysis, MIT PRESS Massachusetts.
Jakobson, R., Waugh, L. R. and Taylor, M. (1979). The sound shape of language. Brighton, Harvester Press.
Jassem, V. (1968). Speech analysis and synthesis. Warsaw.
Jespersen, O. (1904). Phonetische Grundfragen. Leipzig, B.G.Tuebener.
Joos, M. (1948). Acoustic phonetics, Supplement to Language. Journal of the linguistic society of America.
Karnickaya, E. G., Mushnikov, V. N., Slepokurova, N. A. and Zhukov, S. J. (1975). Auditory processing of steady-state vowels. Auditory analysis and perception of speech. C. G. M. Fant and M. Tatham. London, Academic Press: 37-53.
Kiefte, M. and Kluender, K. R. (2005). The relative importance of spectral tilt in monophthongs and diphthongs, The Journal of the Acoustical Society of America 117 (3): 1395-1404.

Kingston, J. and Diehl, R. L. (1994). Phonetic Knowledge, Language: Journal of the Linguistic Society of America 70 (3): 419-454.
Klatt, D. (1982). Prediction of perceived phonetic distance from critical-band spectra: a first step. Proceedings of the IEEE International Conference on Speech Acoustic and Signal Processing, Institute of Electrical and Electronics Engineering, New York.
Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer, Journal of the Acoustical Society of America67 (3) 8–16.
Kraus, N. (1992). Mismatch negativity event-related potential elicited by speech stimuli, Ear and hearing 13 (3): 159-164.
Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not, Perception and Psychophysiology 50 (2): 93-107.
Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception, Journal of Phonetics 21: 125-139.
Kuznetsov, V. and Ott, A. (1987). A spectral properties of Russian stressed vowels in the context of palatalized and non-palatalized consonants. Proceedings XI-th Congress of Phonetic Sciences, Tallinn: 117-120.
Kuznetsov, V. and Ott, A. (2001). Spectral dynamics and classification of Russian vowels. XI Session of Russian Acoustic Society, Moscow: 439–444.
Lacerda, F. (1995). The perceptual magnet effect: an emergent consequence of exemplar-based phonetic memory, Proceedings of International Congress of Phonetic Sciences 2: 140-147.
Laine, U. (1989). Studies on Modelling of Vocal Tract with Applications to speech synthesis. Otaniemi, Helsinki University of Technology, Faculty of Electrical Engineering, Acoustics Laboratory, TKK.
Lammela, J. (2004). Statistical analysis of the vowel identification using general linear mixed models. Studies on speech communication. T. J. and M. Peltola. Turku: 27 - 42.
Liberman, A. M., S., C. F., Shankweiler, D. P. and Studdert-Kennedy, M. (1967). Perception of the speech code, Psychological Review 74 (6): 431-461.
Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast, Language 48 (4).
Lindblom, B. (2000). Developmental origins of adult phonology: the interplay between phonetic emergents and the evolutionary adaptations of sound patterns, Phonetica 57 (2-4): 297-314.
Lotman, M. (2002). Atomistic versus holistic semiotics: 15p.
Macmillan, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. Categorical perception: The

groundwork of cognition. S. Harnad. New York, NY, Cambridge University Press: 53-85.
Maddieson, I. and Disner, S. F. (1984). Patterns of sounds. Cambridge, Cambridge U.P.
Maddox, W. T., Molis, M. and Diehl, R. (2002). Generalizing a neurophysiological model of visual categorization of vowels, Perception Psychophysics 64 (5): 585 - 597.
Milenkovic, P. and Forrest, K. (1988). Classification of vowels using spectrum moments, Journal of the Acoustical Socety of America 83 (s1): s67.
Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel, Journal of the Acoustical Society of America85 (5): 2114-34.
Moore, B. C., Peters, R. W. and Glasberg, B. R. (1990). Auditory filter shapes at low center frequencies, Journal of the Acoustical Society of America 88 (1): 132-40.
Määttä, T. (1983). Hur finskspråkiga uppfattar svenskans vokaler, Universitetet i Umeå.
Nearey, T. and Kiefte, M. (2003). Comparison of several proposed perceptual representations of vowel spectra. 15th ICPhS, Barcelona, Spain, Universitat Autònoma de Barcelona: 1005–1008.
Nearey, T. M. (1989). Static, dynamic and relational properties in vowel perception, Journal of the Acoustical Society of America 88 (1): 3241-3254.
Nearey, T. M. (1997). Speech perception as pattern recognition, Journal of the Acoustical Society of America101 (6): 3241-3254.
Nenonen, S., Shestakova, A., Houtilainen, M. and Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing, Cognitive Brain Research 16 (3): 492-495.
Niemi, M., Laaksonen, J.-P., Aaltonen, O. and Happonen, R.-P. (2004). Effects of transitory lingual nerve impairment on speech: an acoustic study of diphthong sounds, Journal of Oral and Maxillofacial Surgery 62 (1): 44-51.
Niemi, M., Laaksonen, J.-P., Vähätalo, K., Tuomainen, J., Aaltonen, O. and Happonen, R.-P. (2002). Effects of transitory lingual nerve impairment on speech: An acoustic study of vowel sounds, Journal of Oral and Maxillofacial Surgery 60 (6): 647-652.
Niemi, P. and Aaltonen, O. (1986). Tutkimusmenetelmät puheen tunnistuksen mallien synnyttäjinä, Virittäjä 90: 203-211.
Nord, L. and Svantelius, E. (1979). Analysis and prediction of difference limen data for formant frequencies, TMH-QPSR 3-4: 60-70.

Näätänen, R. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses, Nature 385 (30): 432-434.
Näätänen, R. and Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience, Psychological Bulletin 125 (6): 826-859.
Ohl, F. W. and Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction, PNAS 94 (17): 9440-9444.
Okada, H. (1999). Japanese. Handbook of the International Phonetic Association: A guide to usage of the International Phonetic Alphabet. Cambridge, Cambridge University Press: 117-119.
Oppenheim, A., Schafer, R. W. and Buck, J. R. (1999). Discrete-Time Signal Processing. Englewood-Cliffs, NJ, Perentice-Hall.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L. and Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold, Journal of the Acoustical Society of America 72: 1788-1789.
Peirce, C. S., Kloesel, C. J. W. and Houser, N. (1992). The essential Peirce : selected philosophical writings. Bloomington, Indiana University Press.
Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of vowels, Journal of the Acoustical Society of America 24 (2): 175-184.
Pisoni, D. B. and Luce, P. A. (1987). Acoustic-phonetic representations in word recognition, Cognition 25 (1-2): 21-52.
Polka, L. and Bohn, O.-S. (2003). Asymmetries in vowel perception, Speech Communication 41 (1): 221-231.
Pols, L. C., Van der Kamp, L. J. and Plomp, R. (1969). Perceptual and physical spece of vowel sounds, Journal of the Acoustical Society of America (2) 46 (2).
Raimo, I., Savela, J. and Aaltonen, O. (2003). Turku Vowel Test. Fonetiikan päivät, Akustikan ja äänenkäsittelytekniikan laboratorio,, TKK, Otaniemi: 45-52.
Raimo, I., Savela, J. and Aaltonen, O. (2005). Vokaalit testissä. Puheen Salaisuudet. A. Iivonen: 171-181.
Raimo, I., Savela, J., Launonen, A., Kärki, T., Mattila, M., Uusipaikka, E. and Aaltonen, O. (2002). Multilingual Vowel Perception. ISCA Workshop on Temporal Integration in Perception of Speech, 8.- 10. April 2003 France: 86.
Ravila, P. (1967). Totuus ja metodi : kielitieteellisiä esseitä. Porvoo, Söderström.
Rédei, K. (1978). Chrestomatica Syrjaenica. Budapest, Tankönyvkiadó.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S. and Lang, J. M. (1994). On the perceptual organization of speech, Psychological Review 101 (1): 129-56.
Repp, B. H. and Crowder, R. G. (1990). Stimulus order effrects in vowel discrimination,), Journal of the Acoustical Society of America88 (5): 2080-2090.
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 (1): 129-145.
Rodman, R., McAllister, D., Bitzer, D., Cepeda, L. and Abbitt, P. (2002). Forensic speaker identification, Forensic Linguistics 9 (1): 22-43.
Rosner, B. S. and Pickering, J. B. (1994). Vowel perception and production. Oxford, Oxford Univ. Press.
Sakayori, S., Kitama, T., Chimoto, S., Qin, L. and Sato, Y. (2002). Critical spectral regions for vowel identification, Neuroscience Research 43 (2): 155-162.
Saussure, F. (1919). Course in linguistic general. Lausanne & Paris, Librairie Payot & Cie.
Savela, J. (1999) Komin ja suomen vokaalit – kontrastiivinen tutkimus. Unpublished Master Thesis, Department of Phonetics, University of Turku.
Savela, J. (1999). Tutkimuskomisyrjäänin ja suomen vokaalifoneemien rakenteista, Sanajalka 41.
Savela, J. (2000). On the acoustic nature of Komi Zyrian vowels. Congressus Nonus Internationalis Fenno-Ugristarum, Tartu : 158-161.
Savela, J., Ek, M., Kujala, T., Lang, A. H., Aaltonen, O. and Näätänen, R. (2000). Comparison of two vowel systems by mismatch negativity. Neuroscience 2000 Finland Symposium, Helsinki.
Savela, J., Kleimola, T., Mäkelä , L., Tuomainen, J. and Aaltonen, O. (2003a). Distinktiivisten piirteiden vaikutus vokaalien havaitsemiseen tietoisella ja esitietoisella tasolla. Fonetiikan päivät, Akustikan ja äänenkäsittelytekniikan laboratorio, TKK, Otaniemi: 39-44.
Savela, J., Kleimola, T., Mäkelä, L., Tuomainen, J. and Aaltonen, O. (2003b). The effects of distinctive features on the perception of vowel categories. the15th International Congress of Phonetic Sciences, Barcelona, Spain, Universitat Autònoma de Barcelona: 1000 – 1003.
Savela, J., Kujala, T., Ek, M., Tuomainen, J., Aaltonen, O. and Näätänen, R. (2001). Suomen ja komin vokaalit esitietoisella ja tietoisella tasolla. 21. Fonetiikan päivät, Turku, Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja: 121-127.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O. and Näätänen, R. (2003). The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages, Cognitive Brain Research 16 (2): 250-256.
Savela, J., Kujala, T., Tuomainen, J., Ek, M., Altonen, O. and Näätänen, R. (2002). Comparison of two vowel systems by the MMN and reaction times. Temporal intergration in the Perception of Speech, Aix-en-Provence, Cambridge University Press, UK.: 89.
Savela, J., Ojala, S., Aaltonen, O. and Salakoski, T. (2007). Role of different spectral attributes in vowel categorisation: the case of Udmurt. Nodalida 2007, Tarttu, University of Tarttu.
Savela, J. and Pikkanen, O. (2005). Role of the higher formants in vowel identification. Studies in Speech perception. T. J. Turku, Department of Finnish and General linguistics: 15-26.
Savela, J., Pikkanen, O., Raimo, I., Uusipaikka, E. and Aaltonen, O. (2004). Stability of vowel perception - results of the Turku Vowel test. Fonetiikan päivät 2004 - The phonetic symposium, Oulu. 30-31.
Schouten, B., Gerrits, E. and van Hessen, A. (2003). The end of categorical perception as we know it, Speech Communication 41 (1): 71-80.
Schwartz, J.-L., Boe, L.-J., Vallee, N. and Abry, C. (1997a). The Dispersion-Focalization Theory of Vowel Systems, Journal of Phonetics 25 (3): 255-283.
Schwartz, J.-L., Boe, L.-J., Vallee, N. and Abry, C. (1997b). Major Trends in Vowel System Inventories, Journal of Phonetics 25 (3): 233-253.
Sharma, A. and Dorman, M. F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential, Journal of the Acoustical Society of America 104 (1): 511-517.
Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects, Psychometrika 39: 373 - 421.
Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R. J. and Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study, Neuroreport 13 (14): 1813-1816.
Sievers, E. (1881). Grundzüge der Phonetik. Leipzig, Breitkopf & Hartel.
Sluijter, A. M. and van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress, Journal of the Acoustical Society of America 100 (4 Pt 1): 2471-2485.
Stevens, K. N. (1980). Acoustic correlates of some phonetic categories, Journal of the Acoustical Society of America 68 (3): 836-842.
Stevens, K. N. (1989). On the Quantal Nature of Speech, Journal of Phonetics 17 (1-2).

Stevens, S. S. and Volkmann, J. (1940). The relation of pitch to frequency: a revised scale, American Journal of Psychology 53: 329-353.
Strange, W. (1989). Evolving theories of vowel perception, Journal of the Acoustical Society of America 85 (5): 2081- 2087.
Studdert-Kennedy, M. (1974). Perception of speech, Current trends in linguistics 12: 2349-2385.
Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P. and Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts, Hearing Research 190 (1-2): 128-140.
Thyer, N., Hickson, L. and Dodd, B. (2000). The perceptual magnet effect in Australian English vowels, Perception and Psychophysics 62 (1): 1-20.
Tobin, Y. (1990). Semiotics and linguistics. Harlow, Longman.
Traunmüller, H. and Lacerda, F. (1987). Perceptual relativity in identification of two-formant vowels, Speech Comm 6:
Vehkavaara, T. (2003). Natural self-interest, interactive representation, and the emergence of objects and Umwelt: An outline of basic semiotic concepts for biosemiotics: 547-587.
Versnel, H. and Shamma, S. A. (1998). Spectral-ripple representation of steady-state vowels in primary auditory cortex, The Journal of the Acoustical Society of America 103 (5): 2502-2514.
Winkler, I., Reinikainen, K. and Näätänen, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans, Perception and Psychophysiology 53 (4): 443-449.
Zwicker, E. (1961). Subvision of the audible frequency range into critical bandwidths (Frequenzgruppen), Journal of the Acoustical Society of America 33: 248.

Turku Centre for Computer Science
TUCS Dissertations

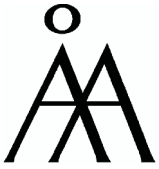
86. **Sanna Ranto**, Identifying and Locating-Dominating Codes in Binary Hamming Spaces
87. **Tuomas Hakkarainen**, On the Computation of the Class Numbers of Real Abelian Fields
88. **Elena Czeizler**, Intricacies of Word Equations
89. **Marcus Alanen**, A Metamodeling Framework for Software Engineering
90. **Filip Ginter**, Towards Information Extraction in the Biomedical Domain: Methods and Resources
91. **Jarkko Paavola**, Signature Ensembles and Receiver Structures for Oversaturated Synchronous DS-CDMA Systems
92. **Arho Virkki**, The Human Respiratory System: Modelling, Analysis and Control
93. **Olli Luoma**, Efficient Methods for Storing and Querying XML Data with Relational Databases
94. **Dubravka Ilić**, Formal Reasoning about Dependability in Model-Driven Development
95. **Kim Solin**, Abstract Algebra of Program Refinement
96. **Tomi Westerlund**, Time Aware Modelling and Analysis of Systems-on-Chip
97. **Kalle Saari**, On the Frequency and Periodicity of Infinite Words
98. **Tomi Kärki**, Similarity Relations on Words: Relational Codes and Periods
99. **Markus M. Mäkelä**, Essays on Software Product Development: A Strategic Management Viewpoint
100. **Roope Vehkalahti**, Class Field Theoretic Methods in the Design of Lattice Signal Constellations
101. **Anne-Maria Ernvall-Hytönen**, On Short Exponential Sums Involving Fourier Coefficients of Holomorphic Cusp Forms
102. **Chang Li**, Parallelism and Complexity in Gene Assembly
103. **Tapio Pahikkala**, New Kernel Functions and Learning Methods for Text and Data Mining
104. **Denis Shestakov**, Search Interfaces on the Web: Querying and Characterizing
105. **Sampo Pyysalo**, A Dependency Parsing Approach to Biomedical Text Mining
106. **Anna Sell**, Mobile Digital Calendars in Knowledge Work
107. **Dorina Marghescu**, Evaluating Multidimensional Visualization Techniques in Data Mining Tasks
108. **Tero Säntti**, A Co-Processor Approach for Efficient Java Execution in Embedded Systems
109. **Kari Salonen**, Setup Optimization in High-Mix Surface Mount PCB Assembly
110. **Pontus Boström**, Formal Design and Verification of Systems Using Domain-Specific Languages
111. **Camilla J. Hollanti**, Order-Theoretic Mehtods for Space-Time Coding: Symmetric and Asymmetric Designs
112. **Heidi Himmanen**, On Transmission System Design for Wireless Broadcasting
113. **Sébastien Lafond**, Simulation of Embedded Systems for Energy Consumption Estimation
114. **Evgeni Tsivtsivadze**, Learning Preferences with Kernel-Based Methods
115. **Petri Salmela**, On Commutation and Conjugacy of Rational Languages and the Fixed Point Method
116. **Siamak Taati**, Conservation Laws in Cellular Automata
117. **Vladimir Rogojin**, Gene Assembly in Stichotrichous Ciliates: Elementary Operations, Parallelism and Computation
118. **Alexey Dudkov**, Chip and Signature Interleaving in DS CDMA Systems
119. **Janne Savela**, Role of Selected Spectral Attributes in the Perception of Synthetic Vowels

Turku
Centre *for*
Computer
Science

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics



Åbo Akademi University

- Department of Information Technologies

Turku School of Economics

- Institute of Information Systems Sciences

ISBN 978-952-12-2308-2
ISSN 1239-1883

Janne Savela

Role of Selected Spectral Attributes in the Perception of Synthetic Vowels

Janne Savela

Role of Selected Spectral Attributes in the Perception of Synthetic Vowels