Role of Different Spectral Attributes in Vowel Categorization: the Case of Udmurt

Janne Savela
Department of Information Technology
University of Turku
janne.savela@utu.fi

Stina Ojala
Department of Information Technology
University of Turku
stina.ojala@utu.fi

Olli Aaltonen
Department of Phonetics
University of Turku
olli.aaltonen@utu.fi

Tapio Salakoski
Department of Information Technology
University of Turku
tapio.salakoski@utu.fi

Abstract

The present study examines the difference between categorization and goodness ratings in Udmurt (Finno-Ugric language) using different sets of spectral attributes. The tendency to have two areas of /u/ vowels was observed in languages with unrounded non-front closed vowels during the TURVOTES project. Our study explores whether this can be due to the different acoustic attributes used in the identification and goodness ratings of these vowels. The identification and goodness-rating of Udmurt close vowels confirmed the observation. The model using only formants was not significant for identification data, but did explain the goodness rating data. The spectral moments explained both identification and goodness ratings.

1 Introduction

The vowel perception is a process in which the auditory pattern induced by the perceived stimulus is categorized with the stored auditory patterns of similar stimuli. The basic question in speech perception studies is to understand the role of the different spectral attributes within these processes and the actual similarity of two sounds within a chosen scale. This is crucial in the interpretation of the patterns observed in human performance based on these similarities (e.g. identification and discrimination and linguistic use of the same physical stimuli). The crucial question is: what type of similarity the listener use, if asked to identify the stimuli or to estimate its goodness. The knowledge on these features makes possible to predict, how people would perceive the stimuli they have not heard before, and consequently categorization in accordance with their linguistic knowledge about the speech sounds.

The long time problem of the vowel identification studies has been the discrepancy between the whole spectral attributes and formants that give different information about the physical similarity of the stimulus. According to Rosner (1994) the discrepancy may be related to the difference between general acoustic distance and the phonetic distance between the same vowel pair (also Granstrom & Klatt, 1979). The general assumption has been that the formant hypothesis, formant peak picking, is enough to explain the attentive vowel identification responses. Savela et al. (2003) showed that the formant peak picking and the general acoustic distance can be related to the different types of perceptual processes: automatic and post-perceptual. The automatic process buffers the fea-

Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek and Mare Koit (Eds.)
NODALIDA 2007 Conference Proceedings, pp. 384–388

tures of the stimuli, using intuitive knowledge on the important information within the stimulus. It was shown (ibidem) that the formant information can be considered as an intuitive indexical knowledge on the category and the whole spectral information was shown to be post-perceptual language independent general information about the similarity of two stimuli. In extreme cases the general auditory knowledge and the formant based linear knowledge can result non-continuous areas within the vowel space.

The aim of the present study is to use comparison of the identification and goodness rating data in order to show the role of formant based intuitive (indexical) similarity and the whole spectral based auditory similarity between the synthetic stimuli, using familiarity based indexical (goodness ratings) and analytical symbolic strategies. The idea is to identify the attributes sufficient to explain the identification and goodness ratings adequately.

First two models are tested, in the first model the efficiency of two formant model is tested and in the second model with the spectral moments added is tested. The aim of the present study is to show that vowel identification and goodness rating are based on different spectral attributes in terms of vowel space. The chosen vowels are Udmurt /ɨ/ (unrounded close central vowel) and /u/ (rounded close back vowel) which both cover the same areas of the F1-F2 vowel space. This area of the perceptual vowel space is critical, because it has been revealed as the most challenging area of the vowel space in using of the formant based models. In this particular area of the vowel space the F1, F2 and F3 are distinct and no trivial fusion of the formants can be observed.

2 The TURVOTES data for Udmurt

2.1 Methods

Stimuli

The test consisted of synthetic vowels which covered the entire vowel space except for diphthongs and nasal vowels. The stimuli were synthesized with Klatt parallel synthesizer (Klatt 1980). The vowel space was created by varying F1 from 250 to 800 Hz with steps of 30 mel and F2 from 600 to 2800 Hz with steps of 50 mel. F3 is 2500 Hz as long as F2 is 2000 Hz or below and higher by 200 mel when F2 is above 2000 Hz. The duration of the vowel stimuli was 350 ms and their pitch rose first from 100 Hz to 120 Hz (until 120 ms) and fell then to 80 Hz during the rest of the stimulus. The amplitude of the formants were not damped which can lead to the higher amplitudes of higher formants.

Procedure

There were six subjects (mean age 28,5). They were asked to name the vowel they heard (using Udmurt orthography) and to evaluate the goodness of the vowel (grades 1 – 7). The test for whole database took about 45 minutes. The stimuli were delivered through headphones in a sound hampered room at the Turku University Language centre.

2.2 Results

The identification data for the Udmurt data is drawn in Figure 1.

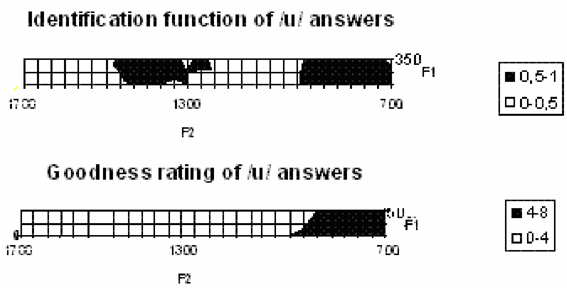

Figure 1. The identification of the non front back vowels in Udmurt. The goodness is described using 1 – 7 goodness scale (two steps) and the identification is described on percent of answers (over 50 % categorization is presented with black color).

The results were clear. In categorization data the area of /u/ was divided into the two different areas whereas the goodness rating data showed only one area of high goodness. Our finding is against the assumption of existence of linear prototype-based categorization of the vowel stimuli. This is because identification of the category can not be explained by two lowest formants of prototypical exemplars.

3 Statistical evaluation of the Udmurt data

3.1 Methods

In order to study the similarity between prototypical areas and vowel boundaries, two binominal regression models were made. The stimuli that the subjects had perceived as /u/ or /ɨ/ were chosen to analysis. Four models were tested in binominal logistic regression mode. In two of them the stimulus category was the dependent parameter and in two of the models the stimulus goodness served as the dependent parameter respectively. The independent parameters were the formants or the formants with spectral moments respectively.

The spectral moments in the signals were analyzed using the PRAAT speech analysis program. First the Fourier analysis of the vowels was made for all stimuli, using sampling frequency 11024 Hz. The measuring of the spectral moments was based on the power spectrum in which the magnitudes of the spectral components are squared. It has been used by Forrest (1988) in measuring of the fricatives. The centre of gravity describes the average frequency of the spectral components in signal (Figure 2).

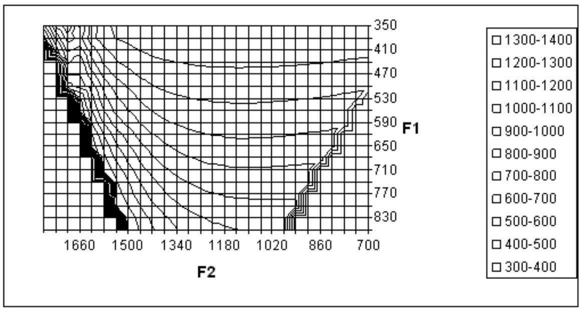

Figure 2. Center of gravity in power spectrum of the synthetic vowel in mel scale plotted against the formant values. The analyzed area is marked with boundaries.

The standard deviation describes the mean difference between the spectral components and the centre of gravity. It is a square root of the second central moment of the spectrum. It describes how much the spectral components differ from the centre of the gravity (Figure 3).

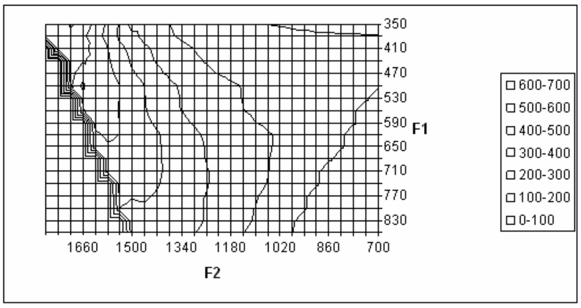

Figure 3. Standard deviation in the power spectrum of the synthetic vowel in mel scale plotted against the formant values. The analyzed area is marked with boundaries.

The (normalized) skewness tells the asymmetry in the shape of the spectrum between the lower and higher areas in the vowel spectrum. It is measured by dividing the third spectral moment (non-normalized skewness) with the 1.5 power of the second spectral moment Figure 4.

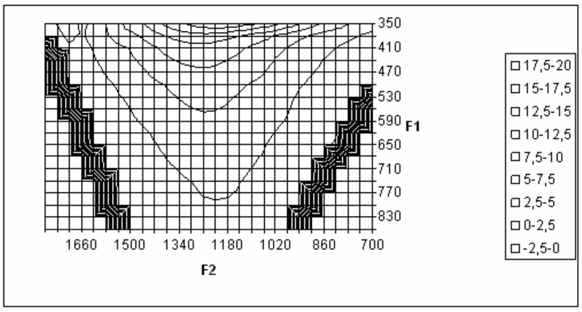

Figure 4. Skewness in the power spectrum of the synthetic vowels plotted against the formant values. The analyzed area is marked with boundaries.

The (normalized) kurtosis describes how much the spectrum differs from the Gaussian distribution (Figure 5).

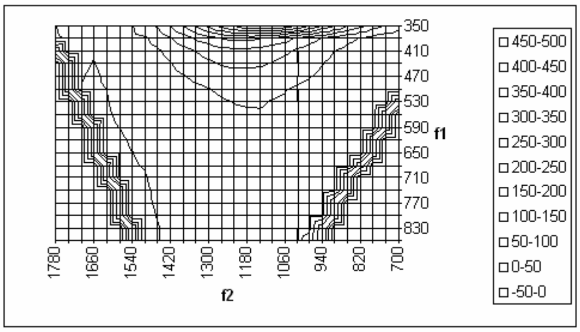

Figure 5. Kurtosis in the power spectrum of the synthetic vowels plotted against the formant values. The analyzed area is marked with boundaries.

3.2 Results

The results of the study are shown in Table 1.

			VOWEL NAME /ɨ - u/			VOWEL GOODNESS/ɨ - u/		
	Chi square	DF	sig.	-2 log	Chi square	DF	sig.	-2 log
F1F2	4,444	2	0,108	424,287	115,488	2	,000	346,146
F1F2 + SPEC	113,799	6	0,00	314,932	143,548	6	0,00	318,086

Table 1. The strength of models in the Udmurt /ɨ-u/ identification.

The first column shows the chi-square value of the model that tells the general fit of the variables in the model (omnibus test on the variables within the model). The next column (DF) tells the degrees of freedom in model and the third column (sig.) tells omnibus significance of the model. In fourth column (-2log likelihood) presents a value that gives the deviation of the likelihood function of the logistic model and makes possible to compare the fits of the models.

The results show that the identification of vowels was not predicted by formants (insignificant) only whereas the goodness ratings could be evaluated using formant values. The effect of the difference between goodness rating models was smaller, although significant. However looking at the individual formants gives a clearer picture (Table 2).

	VOWEL NAME /ɨ - u/			VOWEL GOODNESS/ɨ - u/		
	Chi-square	DF	P	Wald	DF	p
F1	,803	1	,370	10,019	1	0,002
F2	13,194	1	,000	17,829	1	0,000
COG	3,148	1	,076	5,366	1	0,21
STD	7,897	1	,005	5,089	1	0,24
SKEWNESS	18,618	1	0,00	6,540	1	0,011
KURTOSIS	16,748	1	0,00	17,693	1	0,00

Table 2. The statistical significance of different spectral attributes in identification and goodness rating of Udmurt vowels in models with spectral moments.

In goodness ratings F1 was more important than in categorization that was more the importance of skewness and kurtosis was more significant. The results showed that the identification and goodness ratings were based on different acoustic criteria although many features had similar significance.

4 Discussion and Conclusion

The present study was designed to show that the goodness ratings and identification of the vowels use different spectral attributes in the vowel space. The results argued against the use of the similar model of predictive acoustic parameters for goodness rating data and identification data, providing that the formant based similarity does not fit the identification functions of the Udmurt identification data. It was shown that the identification can be based on post-perceptual process different to goodness rating. The plasticity to the general acoustic features can lead to the areas of similarity that are not similar to the proto-typical exemplars of the category, but still categorized as the same.

The general acoustic features can lead to the areas of similarity that are not similar to the prototype. Recently, the understanding of the phonetic experience has exploded. The present study showed that the many questions concerning the area of prototypical vowels and their relationship to the rest of the category can be elaborated using auditory model in which spectral moments are added to the model. In comparison to the Rosner and Pickering's theory in which the vowel categories are always linear in terms of formant space model (1994) which do not explain the identification data. Additionally, a present study provides information about the perception of open back vowels and their relationship to closed back vowels. It can be speculated, that the perceptual similarity between /u/ and /y/ in some languages may be related to the listeners' habits to use spectral moment information, if synthetic stimuli covering large areas of the vowel space is used. In conclusion, to obtain a more comprehensive description of vowel category, the person-independent formant tracking model has to be extended with spectral moment information dependent on person and culture.

References

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): 1090-105.

Carlson, R., Granström, B. and Klatt, D. (1979). Vowel perception: the relative perceptual salience of selected acoustic manipulations, *TMH-QPSR* 3-4.

Ito, M., Tsuchida, J. and Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception, *Journal of the Acoustical Society of America* 110(2): 1141-1149.

Rosner, B. S. and Pickering, J. B. (1994). *Vowel perception and production*. Oxford, Oxford Univ. Press.

Savela, J., Kleimola, T., Mäkelä, L., Tuomainen, J. and Aaltonen, O. (2003). The effects of distinctive features on the perception of vowel categories. *The 15th International Congress of Phonetic Sciences*, Barcelona, Spain, Universitat Autònoma de Barcelona.

Thyer, N., Hickson, L. and Dodd, B. (2000). The perceptual magnet effect in Australian English vowels, *Perception and Psychophysics* 62(1): 1-20.